

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 11:04:49 ; (without alignments)

Post-processing: Minimum Match 0%

Scoring table: BLOSUM62

Title: AAC39336
Perfect score: 5677
Sequence: 1 MESEGETAGPKMKNIVPQTC. IDPFITSRVTPGPDILECGINC 1065

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 9613422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database : PIR 7.3:*

- 1: pir1:*
- 2: pir2:*
- 3: Pir3:*
- 4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	5677	100.0	T52054	cellulose synthase (EC 2.4.1.-) catalytic subunit [validated] - Arabidopsis thaliana
2	3984	70.2	T05351	C;Species: Arabidopsis thaliana (mouse/ear cress)
3	3811	67.1	T0261	C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 31-Dec-2000
4	3801	67.0	F84649	C;Accession: T52054
5	3628	63.9	T08583	R;Arioli, T.; Peng, L.; Betzner, A.S.; Burn, J.; Wittke, W.; Hirth, W.; Camilleri, C. Science, 271, 717-720, 1998
6	3612.5	63.6	T52028	A;Title: Molecular analysis of cellulose biosynthesis in Arabidopsis.
7	3590	63.2	H84604	A;Reference number: 213745; MVID:98111412; PMID:9445479
8	3473.5	61.2	T10797	A;Accession: T52054
9	3317.5	58.4	T04870	A;Status: preliminary; translated from GB/EMBL/DBJ
10	2808	49.5	T10800	A;Molecule type: mRNA
11	2324	40.9	D86157	A;Residues: 1-1065 <ART>
12	2303.5	40.6	T51546	A;Cross-references: EMBL:AF027174; PIDN:AAC39336.1
13	2289.5	40.3	T05630	A;Note: Ath-B
14	2221.5	39.1	T0362	C;Keywords: glycosyltransferase; hexosyltransferase
15	2185.5	39.5	D84746	C;Function:
16	1931	34.0	T20209	A;Description: EC 2.4.1.-; cellulose synthase [validated, MVID:98111412]; involved in
17	1092.5	19.2	T02561	C;Keywords: glycosyltransferase; hexosyltransferase
18	1069	18.8	T02552	C;Function:
19	1043	18.4	T02553	A;Description: EC 2.4.1.-; cellulose synthase [validated, MVID:98111412]; involved in
20	1018.5	17.9	T02560	C;Keywords: glycosyltransferase; hexosyltransferase
21	953.5	16.8	T08918	C;Function:
22	948.5	16.7	T08920	A;Description: EC 2.4.1.-; cellulose synthase [validated, MVID:98111412]; involved in
23	939.5	16.5	T08919	C;Keywords: glycosyltransferase; hexosyltransferase
24	934.5	16.5	E71417	C;Function:
25	849	15.0	T71417	A;Description: EC 2.4.1.-; cellulose synthase [validated, MVID:98111412]; involved in
26	441.5	7.8	A8275	C;Keywords: glycosyltransferase; hexosyltransferase
27	305.5	5.4	T08591	C;Function:
28	301	5.3	T12093	A;Description: EC 2.4.1.-; cellulose synthase [validated, MVID:98111412]; involved in
29	2831.5	5.0	T08592	C;Keywords: glycosyltransferase; hexosyltransferase

ALIGNMENTS

QY	1	MESEGETAGPKMKNIVPQTC	1	SCPOCKTRYKALKGSPAIPGDKEGLADECTVEENYPOKEKISERMIGHLTRKGEM
Db	1	MESEGETAGPKMKNIVPQTC	1	SCPOCKTRYKALKGSPAIPGDKEGLADECTVEENYPOKEKISERMIGHLTRKGEM
QY	61	SCPOCKTRYKALKGSPAIPGDKEGLADECTVEENYPOKEKISERMIGHLTRKGEM	60	1 GEPQYKEVNSNHPLRITSQDTSEFSAASPERVSSTAGGERLPSSDWNQSPNRR 180
Db	61	SCPOCKTRYKALKGSPAIPGDKEGLADECTVEENYPOKEKISERMIGHLTRKGEM	60	121 GEPQYKEVNSNHPLRITSQDTSEFSAASPERVSSTAGGERLPSSDWNQSPNRR 180
QY	181	IVDPVGIGVNAWKERIDGWKKQKERNINGPYSTQAAASERGGVDIDASTDILADEALINDEA	240	181 IVDPVGIGVNAWKERIDGWKKQKERNINGPYSTQAAASERGGVDIDASTDILADEALINDEA 240
Db	181	IVDPVGIGVNAWKERIDGWKKQKERNINGPYSTQAAASERGGVDIDASTDILADEALINDEA	240	181 IVDPVGIGVNAWKERIDGWKKQKERNINGPYSTQAAASERGGVDIDASTDILADEALINDEA 240
QY	241	ROPLSRKVSPSSRNPYRMIMLRVILCFLHWRITNPVNPAALMWYVSCITWFAL	300	301 SWILODPKPWPNRETYLDRLALRDREGEPSQLAVIDFVSTDPLKEPPVLTAVL 360
Db	241	ROPLSRKVSPSSRNPYRMIMLRVILCFLHWRITNPVNPAALMWYVSCITWFAL	300	301 SWILODPKPWPNRETYLDRLALRDREGEPSQLAVIDFVSTDPLKEPPVLTAVL 360
QY	361	SILAVDYPVKVSCVYFDGAGMLSPESLAETSERPARWKWPCKKYSIERRAPEWFAAK	420	361 SILAVDYPVKVSCVYFDGAGMLSPESLAETSERPARWKWPCKKYSIERRAPEWFAAK 420
Db	361	SILAVDYPVKVSCVYFDGAGMLSPESLAETSERPARWKWPCKKYSIERRAPEWFAAK	420	361 SILAVDYPVKVSCVYFDGAGMLSPESLAETSERPARWKWPCKKYSIERRAPEWFAAK 420

QY	421 IDVKKDKVQTSFVKDRRAKREVEEKFIRINALYSKALKCPEEGWVMODGTPNPGNNNGD 480
Db	421 IDYLKDKVQTFSFKDRRAKREVEEKFIRINALYSKALKCPEEGWVMODGTPNPGNNNGD 480
QY	481 HPMIQLTGLONGGLDAGSGNELLRLVYVSREKRGFQHKKAGAMNALYRVSATLNGPF 540
Db	481 HPMIQLTGLONGGLDAGSGNELLRLVYVSREKRGFQHKKAGAMNALYRVSATLNGPF 540
QY	541 ILMUDCHYINNSKALRAMEACFLMDPNLGKQCVYQFFQRFDGIDKDNYANRNTVFDI 600
Db	541 ILMUDCHYINNSKALRAMEACFLMDPNLGKQCVYQFFQRFDGIDKDNYANRNTVFDI 600
QY	601 NLRGLDG1QGPVVVGTCGVFNRTALLYEPPIKVKHKRPSLSSKLCGGSRKNSKAKES 660
Db	601 NLRGLDG1QGPVVVGTCGVFNRTALLYEPPIKVKHKRPSLSSKLCGGSRKNSKAKES 660
QY	661 DKRSGRHMDFDSTPVFVNIDIEGEGVAGFDDIKKALUMQSLSEREGQSAVETVTIME 720
Db	661 DKKRSGRHMDFDSTPVFVNIDIEGEGVAGFDDIKKALUMQSLSEREGQSAVETVTIME 720
QY	721 NGGVPSATPENILKEATHHVISCGEDKSDDWGMETGWIGSVTFDILGKMHARGWISI 780
Db	721 NGGVPSATPENILKEATHHVISCGEDKSDDWGMETGWIGSVTFDILGKMHARGWISI 780
QY	781 YCMPLKPKFGSAPINLSDRLNQYLRWALGSVETLFSRHCPWYNGYERLKFLERAYVN 840
Db	781 YCMPLKPKFGSAPINLSDRLNQYLRWALGSVETLFSRHCPWYNGYERLKFLERAYVN 840
QY	841 TTYPITPSIPLMCTLAUCLFTNQFIPQSNIASTWFLSFLSTPATGILEMRSGV 900
Db	841 TTYPITPSIPLMCTLAUCLFTNQFIPQSNIASTWFLSFLSTPATGILEMRSGV 900
QY	901 GIDEMWRHQFWTIGGVS AHLFAVFOGQIKVLAGIDTNTVSKASDEGDGFABLYLKW 960
Db	901 GIDEMWRHQFWTIGGVS AHLFAVFOGQIKVLAGIDTNTVSKASDEGDGFABLYLKW 960
QY	961 TLLIPIPTTLTINLNGWVAGVSYAINSGYQSMQPLFGKLEFAWTWVIVHLPFLKGMR 1020
Db	961 TLLIPIPTTLTINLNGWVAGVSYAINSGYQSMQPLFGKLEFAWTWVIVHLPFLKGMR 1020
QY	1021 QNRPTIVVWSVLASIFSLNWRIDFPTSRTVIGPDILECGINC 1065
Db	1021 QNRPTIVVWSVLASIFSLNWRIDFPTSRTVIGPDILECGINC 1065
RESULT 2	
TO5351	cellulose synthase (EC 2.4.1.-) catalytic chain RSW1 - <i>Arabidopsis thaliana</i>
N	Alternative names: protein F8B4_110
C	Species: <i>Arabidopsis thaliana</i> (mouse-ear cress)
D	Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
A	Accession: TO5351
R	Bevan, M.; Terry, N.; Ardiles, W.; Buryshaert, C.; Dasseville, R.; De Clerck, R.; Deerves, H.W.; Mayer, K.F.X.; Schueler, C.
B	Submitted to the Protein Sequence Database, February 1999
A	Accession number: 215409
A	Molecule type: DNA
A	Residues: 1-1081 <BEV>
A	Cross-references: EMBL:AL034567
A	Experimental source: cultivar Columbia; BAC clone F8B4
C	Genetics:
A	Gene: RSW1
A	Map position: 4
A	Intron: 27/3; 93/1; 150/3; 189/1; 252/2; 341/2; 456/3; 502/3; 544/3; 615/3; 704/1; 761
A	Note: F8B4_110
C	Keywords: glycosyltransferase; hexosyltransferase; P-loop
Query Match	70.2%; Score 3984; DB 2; Length 1081;
Best Local Similarity	69.7%; Pred. No. 4_6e-307;
Mat Chs	741; Conservative 128; mismatches 36; Gaps 12;

QY 339 DIFVSTVDPLKEPLVTAVTIVLTLAVDYPDVKSYFDDGAMLSFESLAETSEFARK 398
 ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 Db 342 DVFVSTVDMKPEPLVTAVTIVLTLAVDYPDVKACYVSSDDGSAMLTPEALSTAEFSKK 401
 QY 393 WPPCKKYSIEPRAPEWYFAKIDYLKDQVTSFKDRAMKREYEERKIRINALVSKA 458
 ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 Db 402 WPPCKKYSIEPRAPEWYFAKIDYLKDQVTSFKDRAMKREYEERKIRINALVSKA 461
 ||:|||||:|||||:|||||:|||||:
 QY 459 KCPBEGWWDGTPWPGNNTGHDGMIOFLGONGGLDREGNEPLRVSRKPGHQ 518
 ||:|||||:|||||:
 Db 462 KIPEDGWMDGTSWPGNHPRDHGMIOFLGHSGGLDGDNEPLRVSRKPGHQ 521
 ||:|||||:
 QY 519 HKKAGAMNLALVRSAVLTMGPFLILDCDHYINNSKALEAMCEMLMDPNLGKQCYVOPP 578
 ||:|||||:|||||:
 Db 522 HKKAGAMNLALVRSAVLTMGPFLILDCDHYINNSKALEAMCEMLMDPNLGKQCYVOPP 581
 ||:|||||:
 QY 579 QRFQGIDKUDRYANRNTVFFDINURGLDQIOPGYVYVGCCFVNFTALYGEPPRKVKHK 638
 ||:|||||:
 Db 582 QRFQGIDKUDRYANRNTVFFDINURGLDQIOPGYVYVGCCFVNQRALYGDPLTEEDLE 641
 ||:|||||:
 QY 639 PSLSLKCGSRKNSKA-KESDKKSRRHTDSTVPUFNLDIEEGEAGGFDDEKL 697
 ||:|||||:
 Db 642 PNLTVKSCGSGRKGSKRIPNYEDNRSKRSNDNPVNPLMNEDEDDEB-GDEDEMALL 699
 ||:|||||:
 QY 698 MSQMSLEKRFQGOSAVFVASTLMEENGVPSPATPENLKEIAHVISCGYEDKSDWGMETCW 757
 ||:|||||:
 Db 700 VSQRKLEKPRKGQSPVETTAFTEMEOGLRSPTNPLTLIKEIAHVISCGYEAKTDWGEIGW 759
 ||:|||||:
 QY 758 IYGSYTEDITGFKHMHAROWRSYICMPKLPFKGKASAPINLSDRINQVIRWALSVETIFS 817
 ||:|||||:
 Db 760 IYGSYTEDITGFKHMHAROWRSYICMPKLPFKGKASAPINLSDRINQVIRWALSVETIFS 819
 ||:|||||:
 QY 818 RHCPIWYGNGRLFELRAYERVNTIYTPTSIPLIMCILAYCLFLTNQFIIQPSNTAS 877
 ||:|||||:
 Db 820 RHCPIWYGNGRLKLERAYINTIVYPTSIPLIMCILAYCLFLTNQFIIQPSNTAS 879
 ||:|||||:
 QY 873 IWFLSLELSLFATGCTLEMWSGCVDEMRNQEFWIGVS AHLFAVFOGICLVLAGIDT 937
 ||:|||||:
 Db 880 LCFMLFASLYASAILELKWDVALEDWTRNEOFWVIGGTSAHHLFAVFOGLKLVFAGIDT 939
 ||:|||||:
 QY 938 NFTVTKSASDEDGDPFAELYLKWKWTLLIPPTTLIIVLNWAGVSYAINGQSWGPF 997
 ||:|||||:
 Db 940 NFTVTKSASDEDGDPFAELYLKWKWTLLIPPTTLIIVLNWAGVSYAINGQSWGPF 999
 ||:|||||:
 QY 998 GKLFPAFWWVTHLPPFLKGIMGRNRTPTIVVVWSVLASIFSLWVRLDPFTSRVTG 1055
 ||:|||||:
 Db 1000 GKLFLFWWWVHLPPFLKGIMGRNRTPTIVVVWSVLASIFSLWVRLDPFTSRVTG 1056
 ||:|||||:
 RESULT 5
 T08533 cellulose synthase (EC 2.4.1.-) catalytic chain - Arabidopsis thaliana
 N;Alternate names: protein T22F8 250
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Accession: T08533; T09014
 R;Sevan, M.; Zimmermann, W.; Gruneisen, A.; Wambutt, R.; Bancroft, I.; Mewes, H. W.; May
 submitted to the Protein Sequence Database; May 1999.
 A;Reference number: Z16442
 A;Accession: T08533
 A;Molecule type: DNA
 A;Residues: 1-1084 <BEV>
 A;Cross-references: EMBL:AL050351; GSDB:GN00062; ATSP:T22F8. 250
 A;Experimental source: cultivar Columbia; BAC clone T22F8.
 R;Arioli, T.; Peng, L.; Bettner, A.S.; Burn, J.; Witte, W.; Herth, C.; Camilleri, C.; H
 Science, 279, 717-720, 1998
 A;Title: Molecular analysis of cellulose biosynthesis in Arabidopsis.
 A;Reference number: Z13745; MUID:98111412; PMID:9445479
 A;Accession: T09014
 A;Status: translated from GB/EMBL/DDJB
 A;Molecule type: mRNA
 A;Residues: 1-1084 <ARI>

C;Keywords: cell wall synthesis; glycosyltransferase; hexosyltransferase
 Query Match 63.9%; Score 3628; DB 2; Length 1084;
 Best Local Similarity 63.0%; Pred. No. Be-279;
 Matches 678; Conservative 165; Mismatches 191; Index 42; Gaps 16;

QY 10 KPMKNIVPOTQCISDNVKGKTVGDRFVACDCSFPCYCPCYERKDGNOQSCPOCTRY 69
 ||:|||||:
 Db 29 RSVOLFSQNCQTCGDIETLTWVSELLEFVACNECAFPVCRPCYERREGNQACPOCKRY 88
 ||:|||||:
 QY 70 KRLKGSPATPGKDEDEGLADEGIVEFENI-POKEKISERMLWGLHTRKGGEEMEQPDKE 128
 ||:|||||:
 Db 89 KRTKGSPPRVDGDEEEFDLDEYEDHGMDPHEAAEALSSRLNTGRG-----GLDSA 142
 ||:|||||:
 Db 129 VSHNHLPRLTSRQDSGERSAASPERLSVSSTTAGKRL--PYSSDVNQSPNRRRTDP- 184
 ||:|||||:
 Db 143 PPGSQQPLITYCDBADMS--DRHALIVPPSTGQYGNVYPAPP-TLOSSAPQRHSVPO 199
 ||:|||||:
 QY 185 ----VGLGVNVAKKERVQDGKMKQEKBNKGPKVQASAE-RGGYDIDASTDILADEL 238
 ||:|||||:
 Db 200 KDTAEYGCGSVAKRDRMVEWKRROGKLEKLOVIKHEGGNNRGSNDDEDD--PDMMPMD 256
 ||:|||||:
 QY 239 EAROPLSRKVSISSRSPNRYPMVIMRLVILCFLHYRITNPVNPNAFLWIVSVCIEWF 298
 ||:|||||:
 Db 257 EGROFLSKRPLPIRSSRSPNRYPMVILCRLAIGLFFFHYRILHPVNDAYGLWLTSVICIEWF 316
 ||:|||||:
 Db 299 ALSMILDQFPKWVNPRTYDLRILARDREGEGPSQLAVIDFVSTVPLKEPPLVNT 358
 ||:|||||:
 Db 317 AVSMILDQFPKWVPIERETYDLRSLRYEKEGRPSGLAPDVPUVSTVPLKEPPLVNT 376
 ||:|||||:
 QY 359 VLSTIAVDPYDVKSYFDDGAMLSFESLATESEFARKWPKKYSIEPAPENFA 418
 ||:|||||:
 Db 377 VLSVIAVDPYDVKACYVSSDGKMLTFEALSDTAFARKWVPPCKKENIEPRAPEWFS 436
 ||:|||||:
 QY 419 AKIDYKLKVQTSFKDRAMKREYEERKIRINALVSKALKCPEEGWQMDFGPWPWGNNT 478
 ||:|||||:
 Db 437 OKMDYLUKNVHPAFVFRERRAMKRDYEEFKVKINAVLATAQKYPEEGWQMDFGPWPWGNNT 496
 ||:|||||:
 QY 479 GDHPCMIOFLGQNGGLDAEGNEPLRVYVSRKPGHQHHKAGAMMALVRSALVING 538
 ||:|||||:
 Db 497 RDHPCMIOFLGHSGVRLDGTGDNLRPLRVYVSRKPGDFHKKRAGAMSLIRVSAVLUS 556
 ||:|||||:
 QY 539 PFTLNDCHYINNSKALEAMCEMLDPMNGKQCYVOPFORDGIDRDRHDSNRRNVFF 598
 ||:|||||:
 Db 557 PYLINVDCDHYINNSKALEAMCEMLDPMNGKQCYVOPFORDGIDRDRHDSNRRNVFF 616
 ||:|||||:
 QY 599 DINURGLDQIOPGVYVGCGVFRALYGEPEPIKKKP----SLSK---LCGCSR 650
 ||:|||||:
 Db 617 DINURGLDQIOPGVYVGCGVFRALYGEPEPIKKKP----SLSK---LCGCSR 673
 ||:|||||:
 QY 651 KKN-SKAKKESDKKKSGRHTDSTVVFNLIDIEEGVEGAGFDEDEKALLMSQMSLEKRGQ 709
 ||:|||||:
 Db 674 KSKSKTAK--DKTKNTKETSKOHA--LENDEGVYVPUVSNVEKSEATOPLKLEKRGQ 728
 ||:|||||:
 QY 710 SAVFVASTMENGVPSPATPENLKEIAHVISCGYERKSDMMEIGLYGSPEDING 769
 ||:|||||:
 Db 729 SPYVVASVQLQNGCPVNPASACPLREIQVSCGKSYEDTING 788
 ||:|||||:
 QY 770 FKMHRGGRSRYCMPKLPFKGKASAPINLSDRINQVLSRHCPIIWYGRNG 829
 ||:|||||:
 Db 789 FKMHRGGRSRYCMPKLPFKGKASAPINLSDRINQVLSRHCPIIWYGRNG 848
 ||:|||||:
 QY 830 LKFELRFAYNTIYTPTSIPLIMCILAYCLFTNQFIIQPSNTASIFSLWIFSLFSLIFA 889
 ||:|||||:
 Db 849 LKMFHRGGRSRYCMPKLPFKGKASAPINLSDRINQVLSRHCPIIWYGRNG 908
 ||:|||||:
 QY 890 TGILEMRMSGVGIDDEWRNEOFWVIGGVS AHLFAVFOGILKVLAGIDTNPVTKSASDED 949

QY 966 PPTTLIIVNLYGWVAGSYAANSGYQSWGPFLGKLFAPFWIVHLYPFLKGMLGRQRRTP 1025
 ||||| :||||| :||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 Db 873 PPPLTLLIVNMVGWVAGFSDALNKYEAWGLPGKVFFSEWLHLYPLKGMLGRQRRTP 932

QY 1026 TIWWMSVLLASIFSLIWNRIDPFTSRVTPGD--CGINC 1065
 ||||| :||||| :||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 Db 933 TIWLLMSVLLASIFSLIWNRPVSTADSTTVSQSCISIDC 974

RESULT 9

T04870 cellulose synthase (EC 2.4.1.-) catalytic chain F28A21-190 - Arabidopsis thaliana
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
 C;Accession: T04870
 R;Bevan, M.; Mueller, M.W.; Muellein, A.; Feiber, R.; Bancroft, I.; Mewes, H.W.; Mayer, A.; Reference number: Z15387
 A;Accession: T04870
 A;Molecule type: DNA
 A;Residues: 1-938 <BREV>
 A;Cross-references: EMBL:AL035526
 A;Experimental source: cultivar Columbia; BAC clone F28A21
 C;Genetics:

A;MAP position: 4
 A;Intron: 60/1; 67/3; 94/1; 149/2; 240/2; 355/3; 401/3; 443/3; 514/3; 662/3; 780/3; 791
 A;Note: F28A21-190
 C;Keywords: glycosyltransferase; hexosyltransferase

Query Match Best Local Similarity 58.4%; Score 3317.5; DB 2; Length 958; Matches 639; Conservative 124; Mismatches 173; Indels 119; Gaps 16; QY 20 COICSDNGKTYWDGFREYACDICSFPVPCYEVERKGNOSCOPCKTRKRLKGSPAI 79
 Db 9 CWCGEIGVNSGEFTVACHCSEPICKACLEYREGRCLRC-----GPAI 57

QY 80 GDKDEGLADECTVEFNYPQEKISERMLGWLHTRKGKEEMGEPOYDKEVSHNLPRLT 139
 Db 58 -----IDENV-FDDVEIKTS 71

QY 140 RQDTSGEFSASPERLVSWSSTAGKRLPYSSDVNOSPNRRFDPVGLGNAYAWKERVDW 199
 Db 72 K-TQSISDGIIHARHISTVSTI-----DSELNDE-----YGNPITWKNRVSEW 111

QY 200 KHOKEKTI----GPVSTQAASERGGYDIDASTDILADEALINDE---ARQPLSKRVS 252
 Db 112 KDKDKKKKKKKDKPKTAQHEAQI----PTQHMEDTPNTESGATDYLWVPIR 167

QY 253 SRINPYRMVIMLRIVLICLFLHYRNPVPAFALWLVVICEWALSWILDQFPKWP 312
 Db 168 TRITSYRIVIINRLILITALFENYRITHPVD SAYGLWLTVCIEIWASWVLQDFKWP 227

QY 313 VRETWLDRLLALRYDREGEPSONLAVIDFVSTWDPIKEPPLYNTAVSLLAVIDYPVDK 372
 Db 228 INRETYIDRLSARFEREGEQSOLAANDFVSTWDPIKEPPLITANTVSLALDYFVDK 287

QY 373 SCYFDGGAMLSFSLEAFSFARKWVPECKYSISTERPAPWYEAKIDYLKDVKYOTSF 432
 Db 288 SCIVSDGGAMLSFSLEAFSFARKWVPECKYSISTERPAPWYEAKIDYLKDVKYOTSF 347

QY 433 VKDRRAKREYEEFKIRINALVSKALKCPECWVQMDGTPWPGNNQGDHPGMIQVQLQN 492
 Db 348 VKERRAMKRDYEFKTRMNALVAKAOKTPECGTMODGTSWPGNNRDRHPGMIOVFLGS 407

QY 493 GOLDAEGLNELPRLVYISREKRGFQHKKAGAMNALVRSAVLTNGPFTLNLDCDMYIN 552
 Db 408 GARDIEGNEPLRVVYISREKRGFQHKKAGAMNALVRSAVLTNAPFLNLDCDMYIN 467

QY 553 SKALREAMCFMLDPNIGKQCVQVFQPFQDFGTDKNDRYANRNTVFDINLRLGQGPV 612
 Db 468 SKAVREAMCFMLDPVQGDVFVOPPFQDFGTDKSDRANRNTVFDINLRLGQGPV 527

QY 613 YWGTCVNRNTALGYEPPIKVKKKPSLLSKLGGSRKKNSKAKKESDKKKSGRHSDT 672
 Db 528 YUGTGTVERROAQGYGSPSKRILPQSSSSCCCLTKK--OPDQESEYKDAAREL 585

QY 673 VPVNLDDEIEGVAGEGFDD-EKALLSMSLEKREGOSAVFAS-TLMENGVPPSAMPE 731
 Db 586 AAIFNLGDLD---NYDEYDRSMLISQTSEFRKTFGLSTVHESTLMENGVPPDSNPs 639

QY 732 NLIKEATHVISCYEDKSDDWMEIGWVYGSWEIDLTGFKMIARGWSIYCPKLPFK 791
 Db 640 TLIKEAIIHVISCYEEKTEWGKEIGWVYGSITEDIAGFKMHCGRWRSIYCMLPRAFK 699

QY 792 SAPINLSDRNLQVRLWALGSVEIFSRHCPWGYN-GRKLTLERFAYVNTIYPTSI 850
 Db 700 SAPINLSDRNLQVRLWALGSVEIFSRHCPWGYN-GRKLTLERFAYVNTIYPTSI 759

QY 851 LLMVCTLLAVCLFTNQFTIPOQTSNIASTWLSLFLSIFATGILEMRNSGVGIDEWNRNEQ 910
 Db 760 LWVACTTIPACILTGKFEIPTLSNLASMFL-----GVSTEDLWRNEQ 802

QY 911 FWVIGGSAHLFAVFGFLKMLAGDNTFTVSKAD-DLEFGELYTVKWTLLIPPRSL 861
 Db 862 LIINLVGWAQGSDALNKYEAWGLPGKVFETAWLHLYPLKGMLGRQRNTPTIVL 921

QY 1031 WSILLASIFSLIWNRIDPFTSRVTPGDPTLCGTC 1065
 Db 922 WSILLASIFSLIWNRPVSKT--DTTLSLNC 953

RESULT 10

T10800 cellulose synthase (EC 2.4.1.-) catalytic chain celA2 - upland cotton (fragment)
 C;Species: Gossypium hirsutum (upland cotton)
 C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 C;Accession: T10800
 R;Pear, J.R.; Kawago, Y.; Schreckengost, W.E.; Delmer D.P.; Stalker, D.M.
 Proc. Natl. Acad. Sci. U.S.A. 93, 12637-12642, 1996
 A;Title: Higher plants contain homologs of the bacterial celA genes encoding the catabolic enzymes involved in cellulose degradation
 A;Reference number: Z17152; MUID:97057296; PMID:8901635
 A;Status: preliminary; translated from cDNA
 A;Molecule type: mRNA
 A;Residues: 1-1685 <PEA>
 A;Cross-references:
 A;Experimental source: strain Acala SJ-2; fiber
 A;Genetics:
 A;Gene: celA2
 C;Function:
 C;Description: involved in synthesis of cellulose
 C;Keywords: glycosyltransferase; hexosyltransferase

Query Match Best Local Similarity 49.5%; Score 2808; DB 2; Length 685; Matches 512; Conservative 77; Mismatches 77; Indels 22; Gaps 6;

QY 397 RKWVPCFKYSISTERPAPWYEAKIDYLKDVKYOTSFYKDRRAKREYEEFKIRINALVSK 456
 1 RRWVPCFKHNVERPAPWYEAKIDYLKDVKYHSFVKERRAMKREYEEFKIRINALVSK 60

QY 457 AUKCPERCEWVWMDGTPPGNNQGDHPGMIOVFLGSONGLDARGNELLRLVYISREKRGF 516
 Db 61 AOKPESCWVWMDGTPPGNNQGDHPGMIOVFLGSONGLDARGNELLRLVYISREKRGF 120

QY 517 OHHKAGAMNALVRSAVLTNGPFTLNLDCDMYINNSKALREAMCFMLDPNIGKQCVQ 576
 Db 121 OHHKAGAMNALVRSAVLTNAPFLNLDCDMYINNSKAMREAMCFMLDPGFKLQVQ 180

QY 577 FQRFQDFGTDKNDRYANRNTVFDINLRLGQGPV 636
 ||||| :||||| :||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

Db 181 FPQRFDGIDDRHRYANRNVFFDINMLGLDGLOGPPVYVTCGVENRQALYGDPPVSKR 240
 Qy :|:|:|:|:|:
 Db 637 KK-----PSLLSKLCGGSSRKNSKAKKE-----SDKKSGRH--TDSTVPENL 678
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 Db 241 PKMTCDCWSWCCCCGGSSRKSKKKGERKGKGLLGLLYGKKGKKGKKNVKKGSAPIVDL 300
 Qy :|:|:|:|:|:
 Db 679 DDIEFGVSGAGFDD-EKALLMSQMSLERFRGOSAVFVASTLMENGCVPPSAIPNLEA 737
 C:Accession: 001 EETEGGLE--GYBELEKSTLMSQNKFEKRFQGOSPVFLASTLMENGGLPEGTNSTLKEA 358
 Qy :|:|:|:|:
 Db 738 INVISCGYEDKSDWGMGEMGWIGVGSVTEILTGKMHARGWSRSTCMPKPAFKGSAPINL 797
 C:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 Db 359 IHWISCGYEEKTGKELGKIGWIGVGSVTEILTGFKWMGRKSYCVPRAFKGSAPINL 418
 Qy :|:|:|:|:
 Db 798 SDRINQVIRWALGSVEILSFRHCPPIWYKINGRKFLERAYVNNTIPTISIPLMVCYL 857
 Qy :|:|:|:|:
 Db 419 SDRLHQVIRWALGSVEIFRSRHCPLWYGGKLUWLERAYINTIVVFFTSIPLACTI 478
 Qy :|:|:|:|:
 Db 858 LAVCLFLTQFIPIPOISNIASTWLSLFSIANGILEKWRWGVGIDEWWRNSOFWVIGV 917
 Qy :|:|:|:|:
 Db 479 PAVCLLTGKFTIPLSNLISLJWVHALFLSIIATGVLELRWLGWSVSIQDWWRNEOFWVIGV 538
 Qy :|:|:|:|:
 Db 918 SAHLFAVFOGIKVLLAGIDTNEFISKADEDGDFAEILKFTTLLIPPTILLIVNG 977
 Qy :|:|:|:|:
 Db 539 SAHLFAVFOGIKVLLAGIDTNEFISKADEDGDFAEILKFTTLLIPPTILLIVNG 597
 Qy :|:|:|:|:
 Db 978 WAGVSYAIANSQWSWGLFLGKLFPAFWWVIVHXPFLKGIMGRNRTPTIVVVWSVILAS 1037
 Qy :|:|:|:|:
 Db 598 WAGVSYAIANSQWSWGLFLGKLFPAFWWVILHXPFLKGIMGRNRTPTIVVVWSVILAS 657
 Qy :|:|:|:|:
 Db 1038 IFSLWVWRDPPFSPRVTRDILSCGINC 1065
 Qy :|:|:|:|:
 Db 658 IFSLWVWRDPPFSPRVTRDILSCGINC 685

RESULT 11

Db 86157 hypothetical protein F22D16_26 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Accession: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
 R:Theologis, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.W.; Chung, M.K.; Conn, D.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Hultzar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; Li, Y.; Liu, X.; Liu, Z.A.; Luro, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, A.; authors: Salzberg, S.L.; Schwartz, J.R.; Shin, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: Ab6141; MUID:21016719; PMID:11130712
 A:Accession: D86157
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1181; <SPro>
 A:Cross-references: GB:AE005172; NID:96056428; PIDN:AAF02892.1; GSPDB:GN00141
 C:Genetics:
 C:Map position: 1

Query Match 40.9%; Score 2324; DB 2; Length 1181;
 Best Local Similarity 44.0%; Pred. No. 24e-175;
 Matches 491; Conservative 165; Mismatches 268; Indels 192; Gaps 29;

Db 18 OTCOR-CSDNVGKTVDFDRFVACDICSPPVCPYCVERKDQNOSCPCOCKTTRYKRKGS 75
 Qy :|:|:|:|:|:
 Db 160 QICWLKGCD--RKVHGC--RCE-CGRICRUCYFDCITSGGNCPCGCKEPRDIND 211
 Qy :|:|:|:|:
 Db 76 PAIPGDKDDEGLADEGTWEPNYQKEKISERMLGWHLIRGKGBEMGEQYDKEYSHNHL 135
 Qy :|:|:|:
 Db 212 PETEEDEDEEARP-----LBP-----MGSSKLKPLS--WV 241
 Qy :|:|:|:
 136 RLTSRQDTSGEFSAASPERLSVSTIAGGKRLPYSSDVNQSPNRRTVDVPGLGNVAKER 195

RESULT 12

Db 242 KSFKAQNQAGDFDHT-----RWLFETK-----GTYGKGNANWPK- 275
 Qy :|:|:|:|:
 Db 196 VDGMKMKQEKNITGPVQASERGGVDASTDILADEALNEAROPLSRVVISPSRI 255
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 Db 276 -DGTG-----GSGGGNGETPPE-----FSBRSKREPLTRKVSVSAAII 314
 Qy :|:|:|:
 Db 256 NPYRMVIMRLVNICLFLHYRITNPVNAFALMVLVICETWFALSWILDQPPKWFFVNR 315
 Qy :|:|:|:
 Db 315 SPYRLALRLVALGALFELTWVRHPRNRAWMWGMGTMCELMFALSMLDQLPKLCVNR 374
 Qy :|:|:|:
 Db 316 ETYLDRLALRYDRBG--EP--SOLAADVIFSTVDPLIKEPLPVLTANVLISLAVDYPVD 370
 Qy :|:|:|:
 Db 375 LTDIGVLKERFESPNLRNPGRKSDLPGIDVFVSTADPEKEPLVATVILSILAVDYPVE 434
 Qy :|:|:|:
 Db 371 KVSCYVFDDGAMLSFEELAISEFARWVWPKKKSYSTEPRAPEWYPAKTDYKDVQ 430
 C:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 Db 435 KLACYLSDDGGALLTFFEAQATASFASTWVPECRKHNEPRMEAYTYGQKRNFLKVKRL 494
 Qy :|:|:|:
 Db 431 SFVKDRRAMKREVEEKIRINAL-----VSKALKC----- 460
 Qy :|:|:|:
 Db 495 DFEVRERYKREYDEEVKVRINSPEAARRSDAYNHWELRAKKKOMEMMMGNPQEVNI 554
 Qy :|:|:|:
 Db 461 -PERGVMQDGTPWPG-----NTQDHPGMIQVFLQGONG----GLDAEON----- 500
 Qy :|:|:|:
 Db 555 VPRAKAW-MSDGSHWGPWSSGETDNRDAGHIIQAMLAAPPNAEVVGAEEADNLDT 613
 Qy :|:|:|:
 Db 501 ---ELPRIVVYVSEKRGKQFHKKAGAMNALVRSVAVLTNEPFILNDCHYINNSKAL 556
 Qy :|:|:|:
 Db 614 DVDIRLPMVYVSEKRGYDHKKAGAMNALVTSALMSNGFILNDCDHIFTYNSAL 673
 Qy :|:|:|:
 Db 557 REAMCFMDPNLKGQKVQVQFQRFGDKNDRAYERNTVFFDINLGLDQ1GQPVVGT 616
 Qy :|:|:|:
 Db 674 REGMCFMLDRG-GDRICVQFOPREGIDPNDRYANTHNTFEDVSMRALDGLQPMVGT 732
 Qy :|:|:|:
 Db 617 GCVNRTALYGYEPPIKVKKKPSLSSLKUGGSRRKNSKAKKESDKKKSGRHTDSTVVF 676
 Qy :|:|:|:
 Db 733 GCIFRRTALYGFSPPRATEHH-----GWGRKISLRRPKAMKKDDEVSLP 783
 Qy :|:|:|:
 Db 677 NLDDIEEGVEGAGFDFDEKALMSQMSLKKERKGOSAVEVAS-----TLMENG-- 722
 Qy :|:|:|:
 Db 784 GEVNEEENDG---DIESLL---LPKRGFNSNSFASIVAEYKOCRLQDQGCKN 834
 Qy :|:|:|:
 Db 835 SRPAGSLAVPREPLDAATVAEAISVISCFYEDTEWGKRVGWIYGSYTEDVWVYGRHMNR 894
 Qy :|:|:|:
 Db 776 GWISIYCMLPKPAFKGSAPINLSDRNLQYLRWALGSMVTEILFSRHCPIWYGYNRKLKLER 835
 Qy :|:|:|:
 Db 895 GWISIYCMLPKDAFGRATPINLDRHOLYRWTGSMFELFESRNAAFP--ATTRMKFLQR 952
 Qy :|:|:|:
 Db 836 FAYVNNTYPTISPLMCTLIACVCTNQFTIPIQNSIATSTWFLSFLSIFATMEL 895
 Qy :|:|:|:
 Db 953 VAYFNGNMVFTSLFELTYCILPAISLFSQFVQVSDITFLYLSTLTCMLSLIEI 1012
 Qy :|:|:|:
 Db 896 RWGSGVGIDEWWRNEQFWVIGGVS AHLFAVFOGLIKVLAGIDTNTFTVSKAS- DEDG- F 952
 Qy :|:|:|:
 Db 1013 KWSGTTLHEWWRNQEFWVIGTSAHAPAVLQGLLKVIAQDVISFTLTSKSSAPEDGDEF 1072
 Qy :|:|:|:
 Db 953 AELYLFKWTLLIPPTILLVNLVWVAGVSVIAINSQYOSWGLPLEGKLFFAWVWVLYP 1012
 Qy :|:|:|:
 Db 1073 ADLVYKNSFMLPPPLTINMVNIAIAYGLARTLYSPFPPQSKLVLGGVFFEWLCHLYP 1132
 Qy :|:|:|:
 Db 1013 FLKGMLGRNRTPTIVVVWSVILASISLWLWVIRD 1048
 Qy :|:|:|:
 Db 1133 FANGLMGRGRVTVTFWSGLSISIISLWLWVYINP 1168
 Qy :|:|:|:
 C:Species: Arabidopsis thaliana (mouse-ear cress)

C; Date: 18-Aug-2000 #sequence_revision 18-Aug-2000

C; Accession: T51546

R; Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Mewes, H.W.

A; Reference number: Z225394

A; Accession: T51546

A; Status: preliminary

A; Molecule type: DNA

A; Residues: 1-1145 <SAT>

A; Cross-references: EMBL:AL1391141

C; Experimental source: cultivar Columbia; BAC clone F2K13

A; Map position: 5

A; Introns: 297/2; 566/3

A; Note: F2K13_60

Query Match Score 40.6%; Length 1145; Best Local Similarity 43.8%; Pred. No. 9.6e-174; Matches 490; Conservative 162; Mismatches 265; Indels 201; Gaps 28;

OY 19 TCQI---CSDNVGKTVGDGRFVACDCSFPVCRCPEYEYERKDGNOSCPQCCKTRYKRKGSP 76

Db 127 SCAIPGCCAKVMSDERRGQDPLPC-CDFKICRCDPFDIAVKTGGGCGCCKEPRY---NT 181

OY 77 AIPGDKDDEGLADEGTVENYPOKEKISERMLGNHLTRKGGEEMEPEQYKEVSINHLP 136

Db 182 HLTDQVDENG-----QQRPMM--LPGGGSKM----- 205

OY 137 LRSQDTSGEFSASPERLS-VSSTIAGKRLPYSSDVNOSPNNRTIVDPV--GIGGVNAW 192

Db 206 -----ERRSMVKSTNSKALMRSGDFDR-NRVLFETGTYGNGNAFW 248

OY 193 KERVGWKKMQEKRNTGPVSTOAASERGGVIDASSTDILADEAL-LNDEARQPLSRKVSP 251

Db 249 -----TQDGEGSGKDGCGDGGMGMEAQDLMSPWRPLRKLKIP 289

OY 252 SSILINPYRAVIMRLVILCFLHYRITNPVNAPAFALWLVSYVICEWFAWSWILDOPPKWF 311

Db 290 AGVISPYRLLIFTRIVVIALFLFTWRVKHONPDAYLWLWGMGSVUCELWFALSWLUDQPLKLC 349

OY 312 PVWRETYDLAIRD--REGEP---SQAADVIFVSTVDPLKEPLVTANTVSLAND 366

Db 350 PIURATDQVLFKEFPTPASNPKSDIPLGFDFVNSTADPEKEPLVTANTVSLIAAE 409

OY 367 YPVDKVSCVVFDDGAMISFESLAAESTEPAARKWWPFCKKYSIEPPAPEWFAAKTDLKD 426

Db 410 YPVEKLSCTYVSDPGGALLTEFAMEAASEANIWPFCCRKAIEPRNPDTSFLSKDPKYN 469

OY 427 KVQTSFVKDRRAMKREYEEFKIRINALVS-----K 456

Db 470 KVSDDFVKDRRRYKREFDEPKVRVNLPSDTSIRRSDAYHAREEIKAMKMQNRDDEPM 529

OY 457 ALKCPPEEKWVMQGTPWPG-----NWYGDHGMCIQVFLG-----QNGCLDAE 498

Db 530 PVIKPKAW-MADGTHWPOTWLTSASDHAKHDGHIQWNLKPRSPDEPLGYVSEFRFLD 588

OY 499 GNE---LPLRVVYSSREKRPQFOHHKKAGAMNALVRSAVLTNGPFTILNDCDHYINNSKL 556

Db 589 DVDIRPLPLVYVSREKRGYDHNKAGAMNALVRSAINSNGPFTLNDCDHYINNSEAL 648

OY 557 REAMCFMDPPLNQSKQCVYOFOPRDGIDKNDRYANRNTVFFDMLRLGUGIQLGYVGT 616

Db 649 REGCMCFMDRG-GDRCLCYQVOPQRREGIDPSDRANHTVFFDVMRNLADGLMGMVYVGT 707

OY 617 GCFNVRTALYGYRPPPIKHKKSLSKCGGSKKN---SKAKKESDKKKSGHDTST 672

Db 708 GCFPRLALYGENP--RSIDFSPCSWCCPSPRKSKNPKENRALRMSD----- 755

OY 673 VPVFNLDIEEGVEGAGFDEDEKALLMSQMSLEKRGROSAYFVAS-----TL 718

OY 756 -----VDEEE--MNLSLVKPFGNSTPLIDSIPVAEFGQRPLADHPA 795

OY 719 MENGGVPPSAT-PENNL-----KERAHVTCGYEDKSDMOMEIGNYGSYTEDLIGKFM 772

Db 796 VKNGRPPGALTIPRELLDASTVABIAIVSCWYEDKTEMRSRIGTYGSTEDVWGYRM 855

OY 773 HARGRWSITCYOMPMLPAFKGSSAPINSLDRNQVLRWALGSEWLSRHRCPWIWYGRKLK 832

Db 856 HNRQWKSVICVTKRDAFRGTAIPINLTDLQLQVLRWATGSVEIFFSRNNALL--ASSKMKI 913

OY 833 LERFAYVNTTYPITSIPLMYCPLAVLFTNQFILPQISNISIWIPLSFLSFATGR 892

Db 914 EFADLYMVWKTSMIPPTITIMVLIAWGSRTIYSSVPPQWSLIGGVSFSFWLAHL 973

OY 893 LEMRWSGVIDEWRNEQFLWVIGGVSAAHFAVFOGILKLAGIDNTFTVSKA--SDEDG 950

Db 974 LEKWSGFSLEEMWRNEQFLWVIGGVSAAHFAVFOGILKLAGIDNTFTVSKA--SDEDG 1033

OY 951 DEABLYLRKWTWTLIIPPTILIVLNUVWVAGVSAINSGYOSWGLFLGKJFCAFWRVHU 1010

Db 1034 EFAADLYMVWKTSMIPPTITIMVLIAWGSRTIYSSVPPQWSLIGGVSFSFWLAHL 1093

OY 1011 YPFELKGLMQRQNRPRTIVWWSVLLASIFSLWVRIDP 1048

Db 1094 YPFAKLGMGRGRGRTPTIVWWSGLVAITISLLWVAINP 1131

RESULT 13

T05646 hypothetical protein F2D010_310 - Arabidopsis thaliana

C; Species: Arabidopsis thaliana (mouse-ear cress)

C; Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999

C; Accession: T05646

R; Bevan, M.; Kutzner, M.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayhew, J.; Reference number: Z15420

A; Accession: T05646

A; Molecule type: DNA

A; Residues: 1-1111 <BEV>

A; Cross-references: EMBL:AL035538

A; Experimental source: cultivar Columbia; BAC clone F2D010

C; Genetics:

A; Map position: 4

A; Introns: 139/2; 675/3

A; Note: F2D010_310

Query Match Score 40.3%; Length 1111; Best Local Similarity 43.1%; Pred. No. 1.2e-172; Matches 481; Conservative 158; Mismatches 256; Indels 221; Gaps 28;

OY 23 CSDNVGKTVGDGRFVACDCSFPVCRCPEYEYERKDGNOSCPQCCKTRYKRKGSPALPGK 82

Db 120 CDGNVNMKDBRGKDMVPC-ECRFKICRCDPMDAQE-TGCLCPGCKEQY-----IGDI 169

OY 83 DEGGL---AEGTGVENYPOKEKISERMLGNHLTRKGGEEMEPEQYDKEVSHNHLPLRTS 141

Db 170 DDPLPDPYSSGALPLPAPGKD-----RGNNNNMS-----MMKR 202

OY 142 DTSGEFSASPERLSVSSTIAGKRLPYSSDVNOSPNNRTIVD--PVGLGVNAW-KERVO 197

Db 203 NQNGEF-----DHNRMWTFETQGTYGNGAAYQDMEY 234

OY 198 GWKKMQEKRNTGPVSTOAASERGGVIDASSTDILADEAL-LNDEARQPLSRKVSPISSRNP 257

Db 235 GDDMDEGMRRGMVETA-----DPKWPRLPSRIPIPAIIS 270

OY 258 YRMVIMRLRVILCFLHYRITNPVNAPAFALWLVSYVICEWFAWSWILDOPPKWFVNRET 317

Db 271 YRLLIVTRFVLCFLTWKTRRNPNDEAIDLWLMSTICELWFGSFNLQDQPKLCPINRST 330

OY 318 YDRLALKYDREGEPEP---SQAADVIFVSTVDPLKEPLVTANTVSLAVDYPDKV 372

Db 331 DLEVLRDKFMPSPSNPTGSRDPLGIDLWVSTADEKEPLVTANTVSLAVDYPVEK 390

OY 373 SCVFDGGAMLSRESLAESEPAARKWVPRCKKYSIEPRAPEWFAAKDYLKDQVTSF 432

Db	391 SCYLSDDGGALLSFEAMAEEASFADLWVPPFCRKHNTPERNPDSYFSLKIDPTKNKSRIDF 450	Best Local Similarity	45.0%	Pred.	No.	2.6e-167;	Matches	469;	Conservative	150;	Mismatches	248;	Indices	175;	Gaps	26;																																																												
Qy	433 VKDRRMKREYEEFKIRINALVS-----KAL-----KCPEE 463	Oy	117 GEEMEPOYOKVEVKSHNHLRFLTSRDTSGEFSASPERSVSSTAGG-----KR 166	Db	451 VKDRRKIREYDFKVRINGLPDSIRRDAFNAREEMKALQKQRESGDPTEPVKVRKA 510	Oy	464 GWWMDGTPWPG-----NNTGDHPGMQLOFGLONGQGLDAEGN-----EL 502	Db	48 GDYSQNDY--INTVLMPTPDNQPSGSSTESKSDANRGGGDGPCKMKGNKLRR 104	Oy	511 TW-MADGTHWPGTWASTRBHSKGHDAGLTLQVMKKPSSDPLIGNSDDKVIDFSDTDRL 569	Db	503 PRLVVSVREKPGFQHHKAGAMNALVURSAVLTINGPFELNDDHYINNSKALREAMC 562	Oy	570 PMFVVSVREKPGYDHNNKAKGAMNALVURSAVLNSNGPFLNLDCDHYINCKAVREGMC 629	Qy	563 LMDPILGKQCVYQFPQRDGDKNDRKYANRNTFFDINERGLDGIQGVYVGTCVFNIR 622	Db	630 MMDRG-GEDICYIQFQPREGIDDSRANNNTFFDGMNRAISLGVOGVQVYVGTMFR 688	Oy	623 TALGYEPPKVKHKPSLISKLGGSRKRNKSKAKKESKKSGRHTDSTVPVNLDDE 682	Qy	689 FALIGFDFP-----NPDKLLE-----KKESE-----T 710	Db	683 EGVSGAGFDEKALIMQSOMSLERKFGQS-----AVFVASTLMEENGV---PPSA- 728	Oy	711 EALTNSDFDPD---LDVYQLPKPGNSTLAESIPIAFQGRPLADHPAVKYGRPPGAL 766	Db	729 -TPENLL---KEAHIVSCGYEDKSDMOMEIGMIGYGSYTEDILGFKMHARGWSIYC 782	Oy	767 RVPDRPLDATVABSSVCSWCYDTEKNGDRVWGYIGSYTEDVYTGYMHRNGWSVC 826	Oy	783 MPKIDAFKASAPINISDRNQVRLVALGSEILPSRHCPTWGYNGRLFPLERAYNT 842	Qy	827 ITKDRSFROSPAPINITDRLHQVLWRATGSVEIFFSRNNAIL- ASKRLKFLQLAYLNWG 884	Db	843 IYPTSIPIJLIMCYLTAVLAVLFNTOTIPoisnisiwiFSLFSLFISAFGTCLEMWSGVI 902	Oy	885 IYPTSIPLFLYCFPAFSLPSQGIVFRTLISIPLVYLIMITICLGLAVLEVWSGIGL 944	Db	903 DEWRNEQFWVIGVSAHLEAVFGQILKLAGIDTNFTSKASDEDG- FAELYFW 960	Oy	945 EEWKRNQWLISLGSITSHLYAWVGQVYKLAGIETISLTLLTSGDDDNEDIADLYIWV 1004	Db	961 TLLIPPTILLIVNVVAGVSAINSQSWPLFGFLFATWVIVILYPEFKGLMR 1020	Qy	1005 SSIMIPPTVIAVMNLIATWAFITIYQAVPQWSKLGAFSTFWLHLYPFKGLMR 1064	Oy	1021 QNRTPTIVVWSVLLASISLWWRIDPTSRVGP 1056	Db	1065 RGKIPPTIVVWAGLIAITISLWWTAINP---NTGP 1096	Qy	733 LLKEAIIHVTSGYEDKSDMGMEIWIYGSYTEDILGFKMHARGWSIYCMPLKUPFKGS 792	Db	744 probable cellulose synthase (imported) - Arabidopsis thaliana (mouse-ear cress)	C;Species: Arabidopsis thaliana (mouse-ear cress)	C;Accession: D84741	C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001	Db	709 TVAERIAVIVSCWYEDTENGDRIGWVYIYGSYTEDVYTGYRMHNKGWSVCITRDAFRGT 768	Qy	793 APINSDRLNQVRLWALGSVEILPSRHCPTWGYNGRLFPLERAYVNTIYPTISIPL 852	Db	769 APINSDRLHQVLWRATGSVEIFFSRNNAIL- ATTRFLKFLQLAYLNWGTFPTSTFLV 826	Qy	853 MYCILAVLAVLFNTOTIPoisnisiwiFSLFSLFISAF---GILEMWGCGIDEMRN 908	Db	827 VVCPFLACLFLPSQKFIVOSL---DIHFLSYLCITVTLTISLLEVWSGIGLEWMNR 882	Qy	909 EQFWWIGVSAHLEAVFGQILKLAGIDTNFTSKASDEDG- FAELYFWTLLIP 966	Db	943 PLTTIVNLYVAVIYGASTRIVSIPPQWGRMLGGIFSLWVLTWMPFKGLMRGKVPT 1002	Qy	967 PTTLIPPTILLIVNVVAGVSAINSQSWPLFGFLFATWVIVILYPEFKGLMRNRPRT 1026	Db	1027 IVVWSVLLASISLWWRIDP 1048

Query Match

39.1%; Score 2221.5; DB 2; Length 1036;

Search completed: June 16, 2003, 11:08:32

Search completed: June 16, 2003, 11:08:32

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 compugen ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 11:08:04 ; Search time 32 Seconds

(without alignments)
3558.288 Million cell updates/sec

Title: AAC39336

Perfect score: 5677

Sequence: 1 MESGETAGPKMKNVQPTC.....IDPFTSRVTPGDILECGINC 1065

Scoring table: BLOSUM62

Gapext 10.0 , Gapext 0.5

Searched: 408643 seqs, 100915682 residues

Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AAs.*

```

1: /rgn2_6/ptodata/2/pubbaa/US08_NEW_PUB.pep:*
2: /rgn2_6/ptodata/2/pubbaa/PCT_NEW_PUB.pep:*
3: /rgn2_6/ptodata/2/pubbaa/US06_NEW_PUB.pep:*
4: /rgn2_6/ptodata/2/pubbaa/US06_PUBCOMB.pep:*
5: /rgn2_6/ptodata/2/pubbaa/US07_NEW_PUB.pep:*
6: /rgn2_6/ptodata/2/pubbaa/US07_PUBCOMB.pep:*
7: /rgn2_6/ptodata/2/pubbaa/US08_PUBCOMB.pep:*
8: /rgn2_6/ptodata/2/pubbaa/US09_NEW_PUB.pep:*
9: /rgn2_6/ptodata/2/pubbaa/US09_PUBCOMB.pep:*
10: /rgn2_6/ptodata/2/pubbaa/US10_NEW_PUB.pep:*
11: /rgn2_6/ptodata/2/pubbaa/US10_PUBCOMB.pep:*
12: /rgn2_6/ptodata/2/pubbaa/US60_NEW_PUB.pep:*
13: /rgn2_6/ptodata/2/pubbaa/US60_PUBCOMB.pep:*
14: /rgn2_6/ptodata/2/pubbaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	5677	100.0	1005	10 US-09-900-237-33
2	4507	79.4	1080	10 US-09-900-237-30
3	4133.5	72.8	1881	10 US-09-888-533-8
4	3999.5	70.5	1091	10 US-09-900-237-26
5	3922	69.1	1148	10 US-09-900-237-4
6	3873.5	68.3	1086	10 US-09-900-237-10
7	3843.5	67.7	1165	10 US-09-900-237-8
8	3804.5	67.0	1039	10 US-09-900-237-14
9	3483.5	61.4	701	10 US-09-900-237-32
10	3473.5	61.2	974	10 US-09-888-533-6
11	3316	56.6	793	10 US-09-900-237-18
12	3003.5	52.9	740	10 US-09-900-237-24
13	2808	49.5	685	10 US-09-900-237-31
14	2803	49.4	685	10 US-09-888-533-7
15	2647.5	46.6	610	10 US-09-900-237-16
16	2420	42.6	506	10 US-09-900-237-20
17	2055.5	36.2	431	10 US-09-900-237-28
18	1800	31.7	720	9 US-10-260-046-18
19	1405	24.7	522	9 US-10-260-046-2

RESULT	1
US-09-900-237-33	; Sequence 33, Appl
	; Sequence 30, Appl
	; Sequence 29, Appl
	; Sequence 28, Appl
	; Sequence 27, Appl
	; Sequence 26, Appl
	; Sequence 25, Appl
	; Sequence 24, Appl
	; Sequence 23, Appl
	; Sequence 22, Appl
	; Sequence 21, Appl
	; Sequence 20, Appl
	; Sequence 19, Appl
	; Sequence 18, Appl
	; Sequence 17, Appl
	; Sequence 16, Appl
	; Sequence 15, Appl
	; Sequence 14, Appl
	; Sequence 13, Appl
	; Sequence 12, Appl
	; Sequence 11, Appl
	; Sequence 10, Appl
	; Sequence 9, Appl
	; Sequence 8, Appl
	; Sequence 7, Appl
	; Sequence 6, Appl
	; Sequence 5, Appl
	; Sequence 4, Appl
	; Sequence 3, Appl
	; Sequence 2, Appl
	; Sequence 1, Appl
	; Sequence 0, Appl

ALIGNMENTS
QY 1 MESGETAGPKMKNVQPTC.....IDPFTSRVTPGDILECGINC 1065
Db 1 MESGETAGPKMKNVQPTC.....IDPFTSRVTPGDILECGINC 1065
QY 61 SCPCKTTRKRLKSPAPGDKBDGLADEGTVFENYPOKEKSERMGMWHLRGKGEM 120
Db 61 SCPCKTTRKRLKSPAPGDKBDGLADEGTVFENYPOKEKSERMGMWHLRGKGEM 120
QY 121 GEPQDKEYSHNHLPRLSQDTSGEFSASPERLVSSTIAGKRLPYSSDNQSPRR 180
Db 121 GEPQDKEYSHNHLPRLSQDTSGEFSASPERLVSSTIAGKRLPYSSDNQSPRR 180
QY 181 IVDPVGIGAWKERVGDWKMKGDKNTGVSTOASERGVDDIDASTDLADEFALNDEA 240

Db 957 NVASLWMSLCIFTSILEMRSVCGIDDMWRNEQFWIGVSSHLERFVQGLKVA 1016
 Qy 934 GIDTNNTVTSKASDEODGDAEFLYFKWMTLIPPTTLIVNUGWAVSYAINGYQSW 993
 Db 1017 GDTSFTVTSKGDD-DEEFSLYTFKWMTLIPPTTLINFIGVAGTSNAINGYESW 1075
 Qy 994 GPLFGKLFFAWVIVHYPFLKGVLGRQNRTPTIVWSILLASIFSLWWRIDPFLAKD 1053
 Db 1076 GPLFGKLFFAWVIVHYPFLKGVLGRQNRTPTIVWSILLASIFSLWWRIDPFLAKD 1135
 Qy 1054 TGPDTIBCGINC 1065
 Db 1136 DGPLLEECGLDC 1147

RESULT 6
 US-09-900-237-10
 ; Sequence 10, Application US/09900237
 ; Patent No. US20020120124A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Allen, Stephen
 ; TITLE OF INVENTION: Plant Cellulose Synthases
 ; FILE REFERENCE: B1170 US CIP
 ; CURRENT APPLICATION NUMBER: US/09/900.237
 ; CURRENT FILING DATE: 2001-07-06
 ; PRIOR FILING DATE: 1998-07-14
 ; PRIOR APPLICATION NUMBER: PCT/US99/15871
 ; PRIOR FILING DATE: 1999-07-13
 ; PRIOR APPLICATION NUMBER: 09/720383
 ; PRIOR FILING DATE: 2000-12-21
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 10
 ; LENGTH: 1086
 ; TYPE: PRT
 ; ORGANISM: Zea mays
 ; US-09-900-237-10

Query Match 68.3% Score 3875.5; DB 10; Length 1086;
 Best Local Similarity 66.7%; Pred. No. 0;
 Matches 728; Conservative 145; Mismatches 161; Indels 57; Gaps 18;

Qy 4 EGETASKPMKNTPQTCQICSDNGKTYVGDRREVACDCSFPVCRPCBEYERKGQNSCP 63
 Db 23 DDPGPKRPPRGCGOQCQICGDBGLPGDPPEVACNCSCAFPCYRCDCBEYERGTQNC 82
 Qy 64 QCKTRKRKLSPATDKDEGLADESTVEFNVP - OKEKTESEMRMLGWHLTREK - GEEM 120
 Db 83 QCKTRKRKLGKQVRIGEDCV-DDLDNEFWGDHSQVSADMLYGHMSGRGGDPN 141
 Qy 121 GEPQVKEVSHHHLPRLTSRQ -- DTSGEFSASPERLVSSTSIIAGGR -- LPISSDN 174
 Db 142 GAPOFO - LNPNVPLITNGQMDPPEQHAYPSFMG --- GGGKTHIHPHY - ADPS 193
 Qy 175 QSPNRRLVDP --- VGLGNPAWKERUDGWKMKQEKNGPVSTQASERGGVUDADSTD 228
 Db 194 LPVOPRSMPSKDLAAGYGSVAKMENWKRQER --- - MHOTGNDGGDD -- 245
 Qy 229 ILADEALNDEARQPJSRKVSPSSRNPYRMYTMLRLEVILCFLHYRITNPYENFAWM 288
 Db 246 -- ADLPJLM-DEARQQLSRKJPLPSSQINPYRMIIIRLWLGFFFHYRVMPYNDPAFLW 302
 Qy 289 LVSWEIWAISLWDOPKPKPNRNTYDYLALRDREGPSQLAADVIFVSTVDP 348
 Db 303 LISVICELWFANSWILDOPKPKPNRNTYDYLALRDREGPSQLAADVIFVSTVDP 362
 Qy 349 KEPPLUTANTVSLAVIDYPVKVSCYFDGAMLSFESLAETSFAKRWVFCCKYSI 408
 Db 363 KEPPLUTANTVSLAVIDYPVKVSCYFDGAMLSFESLAETSFAKRWVFCCKYSI 422
 Qy 409 EPRAPENYFAAKDYLKDVKVQTSFVKDRRAMKREYEEKIRINAVLVAISRAKCPPEGWMO 468

Db 423 EPRAPENYFOQIDYLKDVKVAAVNPFRRKAKKEEFKVRINALVAKQKVUPPEGWMO 482
 Qy 469 DCTPWPGENNTGHPGMQOVFLQGONGGLDAEGNELLPRUVYSSRERPGFOHHKKAGMNL 528
 Db 483 DCTPWPGENNRDPMQVOFLQGOSGGLLDCEGENPLRVYSSRERPGYNNHHKAGMNL 542
 Qy 529 VRVASVLTNPAPLNLDOCHYINNSKAKAREMFLMDONLGKCYVFQRFEGIDK 588
 Db 543 VRVASVLTNPAPLNLDOCHYINNSKAKAREMFLMDONLGKCYVFQRFEGIDK 602
 Qy 589 RYANRNTYFEDINLRGLDGIQGVYVGTGCFENRALTGYEPPKVKQKPSLISKLG- 647
 Db 603 RYANRNVVFFDINMKGLDQGPIVGTGCFVFRQALGYDAP---KTRPP--SRTNC 657
 Qy 648 -----OSRKNSKAKKESDKKK - SGRHTDSVVPVFNLDIEGVEGAGFDBB 694
 Db 658 WPKWCFCCCFGNRKQKTTKPTKEKLLFEKKEENGSPAYAEGE 715
 Qy 695 ALIMSONSLEKRGQSAYFVASTLMEENGQVSPATENLKEATHVSGYEDSDWGE 754
 Db 716 AGIVNOOKLEKKFGQSSVFVSTLLENGTLKSASPASLUKEIAHVSGYEDKTDWGE 775
 Qy 755 GWIYGSYTTEDLTGKMHGWSIYCIPKRYAFKGSAAPLNDSLRLHWRALGSI 814
 Db 776 IGWIYGSYTTEDLTGKMHGWSIYCIPKRYAFKGSAAPLNDSLRLHWRALGSI 835
 Qy 815 LFSRHCPITWGYGNGRLKFLERFAYVNTTYPITSIPLMCTLAUCLETFNQTIQSN 874
 Db 836 FFSNHCPITWGYGGGLKFLERFSYINSIVPWMSIPLAUYCTPAICLITGKFITPELN 895
 Qy 875 LASIWFLSFLSFATGJLLEMRSVGVGIDWWRNQEWNTGIGVSAHLRFVFOGLTKVLG 934
 Db 896 VASLWMSLCIFCATSILEMRMSGVGDWWRNQEWNTGIGVSSHLFVFOGLTKVLG 955
 Qy 935 IDTNFTVTSKASDEGPAEFLYFKWMTLIPPTTLIVNVLVGVAGSYAINGYQSG 994
 Db 956 VDTSETVTSKGDD-DEEFSLYTFKWMTLIPPTTLINFIGVAGVNSAINNGYESW 1014
 Qy 995 PLFGKLFFAWVIVHYPFLKGVLGRQNRTPTIVWSILLASIFSLWWRIDPFLAKD 1074
 Db 1015 PLFGKLFFAWVIVHYPFLKGVLGRQNRTPTIVWSILLASIFSLWWRIDPFLAKD 1074
 Qy 1055 GDPDILEBCGINC 1065
 Db 1075 DGPLLEECGLDC 1085

RESULT 7
 US-09-900-237-8
 ; Sequence 8, Application US/09900237
 ; Patent No. US20020120124A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Allen, Stephen
 ; TITLE OF INVENTION: Plant Cellulose Synthases
 ; FILE REFERENCE: B1170 US CIP
 ; CURRENT APPLICATION NUMBER: US/09/900.237
 ; CURRENT FILING DATE: 2001-07-06
 ; PRIOR APPLICATION NUMBER: 60/092,844
 ; PRIOR FILING DATE: 1998-07-14
 ; PRIOR APPLICATION NUMBER: PCT/US99/15871
 ; PRIOR FILING DATE: 1999-07-13
 ; PRIOR APPLICATION NUMBER: 09/720383
 ; PRIOR FILING DATE: 2000-12-21
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 8
 ; LENGTH: 1165
 ; TYPE: PRT
 ; ORGANISM: Zea mays
 ; US-09-900-237-8

Query Match 67.7% Score 3843.5; DB 10; Length 1165;

Best Local Similarity 66.3%; Pred. No: 0; Mismatches 178; Indels 57; Gaps 17; Matches 727; Conservative 134; Mismatches 178; Indels 57; Gaps 17;

Qy 2 ESGETACKPKMKNIVPQTQICSDNVTGDRFVACDICSFPVCRCYERKDQMS 61
Db 94 DRESGAAGGAARRAEAPCQICDEVGVGFDFGRFVAECAPVCRACYEERREESQQA 153

Qy 62 CPOCKTRYKLKGSPAIQCDKDEGLAD-----EGTVEFNPQEKEKISERMLGWHLT 113
Db 154 CPOCTRYKLKGCPRVGDEEDGVDDLEGERGLQDCAHEIDPQ- YVAEMLRAGMS 211

Qy 114 RGKGEEMBEPQYKEVSHNLPRLTSR---DTSGEFSASAERLSVSSTIAGKR --L 167
Db 212 YGRGGD-AHPGFSPV-- PNVPLTINGQMDPDIPEQHAIWPSYMSGG- GGGKRIHPL 265

Qy 168 PYSDVNQSPNRRIVDP--- VGLGNWAKWKRVDGMKMKOKNTGPVSTOAERGGV 221
Db 266 PF-ADPNLFVQPSRMDPDKLAAYGYSWAKWKRMEGKKQKR---LOIVRSREGG 318

Qy 222 DIDASTDILADEALUNDEARQPLRSRKVSISSRINPYRMVIMRLVILCFLHYRITNPV 281
Db 319 DWPGDD-- -ADPLM-DEARPLSRKVPISSRINPYRMVIVRLVILGFFFHYRMPA 374

Qy 282 PNALFALWIVSVCIEIWFAWSWILDQFQPKWVPIVRETYLDRLAIRYDRIGESEPSOLAANDF 341
Db 375 KDAELWISVICIEIWFAWSWILDQFQPKWVPIVRETYLDRLSLRFDEGGQPSOLAPIDF 434

Qy 342 VSTVDPKKEPPVANTVSLISLUDVPTKVSVYFDGAMLSFESLAETSFARKWP 401
Db 435 VSTVDPKKEPPVANTVSLISLUDVPTKVSVYFDGAMLSFESLAETSFARKWP 494

Qy 402 FCKYSTEPRAPEWYFAAKIDYLKDQVTSFYKDRRAMKREYEFKIRINALVSKALKCP 461
Db 495 FSKKFNTBEPRAPEWYFOOKIDYLKDVKVASFVRERRAKREYEFKIRINALVAKWP 554

Qy 462 EGGVMQDGTPWIGNNITGDPGMQLQVFGLQNGGDAEGNEELPVLVSREREKGQFHKK 521
Db 555 EGGWMQDGSPWIGNNITGDPGMQLQVFGLQNGGDAEGNEELPVLVSREREKGQFHKK 614

Qy 522 AGAMNALVRRSAVLTNGPFITLNIDCDHYTINNSKALREAMCFMLDPNQGKQCVWOPORF 581
Db 615 AGAMNALVRRSAVLTNAAYLLNIDCDHYTINNSKALREAMCFMLDPNQGKQCVWOPORF 674

Qy 582 DGJDKNDRYANRNUFEDINLRGLDGTOPGVYVTGCVFVNRTALGYEPPIKVKHHKPS- 640
Db 675 DGJDKNDRYANRNUFEDINMKLDGLOGPIVYGTGCVFRRQALYGDAPP-- KTKKPPS 731

Qy 641 -----LLSKC - GGSRKNSKAKKESDKKSGRHTDSVPUFNLDIEEGVCG 689
Db 732 RTNCNCWPWKCLSCCSRNRNKKKTPKTEKKRFLRFKAENPSAVALEGEDEGAGA- 790

Qy 690 FDEDEKALLMSQMSLEKRTGQSAVFTLAMENGVPPSAOPENLKEAHVTCGYEDKS 749
Db 791 -DIEKAGINVQKLEKFGQSSVFASITLENGTLKASAPASLKEAHVSCGYEDKT 849

Qy 750 DNGMEIGWIGYGSVETDILGFKHMARGWSRIVCMPKLFPAFKGASPAPINSLNQNLWAL 809
Db 850 DWKGEIGNIYGSTEDILGFKHMCHGWRISYICPKRPAFKGASPAPNLNSDLRQHQLRWA 909

Qy 810 GSVEILSRHCPWYGINGRKLERFVNYNTIPTSPILMCTULLAVLFTNFI 869
Db 910 GSVEIFFSKHCPWYGGGLKFLEFRSYISIVPWHSPILAYCULPATCOLTGKPT 969

Qy 930 KVLAGIDTNFTWIKSKASDDEGDGAELYLFKWTLLIPPTLUVLNUVWAGVSYAING 989
Db 1030 KVFGAGIDTSFTWIKSKAGD- DEESELVTFKWTTILLPTTLLINFIGVAGISNATING 1088

Qy 990 YOSWGPFLGKLFKAFWVWHLYPLKLMGRQNRTPTIVWWVSLASIFSLWVDPF 1049

RESULT 8
US-09-900-237-14
; Sequence 14, Application US/09900237
; Patent No. US20020120124A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen
; TITLE OF INVENTION: Plant Cellulose Synthases
; FILE REFERENCE: BB1170 US CIP
; CURRENT APPLICATION NUMBER: US/09/900,237
; CURRENT FILING DATE: 2001-07-06
; PRIORITY NUMBER: 60/092,844
; PRIORITY NUMBER: 60/092,844
; PRIORITY FILING DATE: 1998-07-14
; PRIORITY APPLICATION NUMBER: PCT/US99/15871
; PRIORITY FILING DATE: 1999-07-13
; PRIORITY APPLICATION NUMBER: 09/720383
; PRIORITY FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 14
; LENGTH: 1039
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (201)
; US-09-900-237-14

Query Match Best Local Similarity 67.0%; Score 3804.5; DB 10; Length 1039;
Matches 725; Conservative 117; Mismatches 160; Indels 75; Gaps 16;

Qy 5 GETAKPKMKNIVPQTQICSDNVTGDRFVACDICSFPVCRCYERKDQGNNSCPO 64
Db 22 GHEPRKALKNLDQVCELCGDDGFLVFACNEGFPVCRCYERKGSHCPCQ 81

Qy 65 CKTRYKRUGSPALPGDKDGLADEGTVEFNPQKE---K1SERMGWHLTRKGKEM 120
Db 82 CKTRYKRUGSPALPGDKDGLADEGTVEFNPQKE---K1SERMGWHLTRKGKEM 139

Qy 121 GEPOQDKEVSHNLPRLTS -- RDTSGEFSASAERLSVSSTIAGRLPSSDVNGSPN 178
Db 140 -----DDDNQSOPTPVIAGGRSPVSGF-----PISSNAYGQDML-SSSLHHRV 184

Qy 179 RRYVDPVGLN--- VAWKERYDWGMKOENKTGPVSTOAERGGVVDIDASTDILAD 232
Db 185 PYPVSEPCSSARWDEKKXDGWWDWL-QDQNGLP-----EPDSDP 227

Qy 233 EALINDEARQPLRSRKVSPFSSRNPYRAVTMRLVILCFLAYRTNEVPAFLWVSV 292
Db 228 AAML-DEARQPLSRKVKPVIASSKINPYRAVIVARLVAFLRFLYRMLMPVHDALGWLTSI 286

Qy 293 ICEIWFASWILDQFQPKWVPIVRETYLDRLAIRYDRIGESEPSOLAANDFVSTVDPKLEPP 352
Db 287 ICEIWFASWILDQFQPKWVPIVRETYLDRLAIRYDRIGESEPSOLAANDFVSTVDPKLEPP 346

Qy 353 LVTANTVSLAVDVPDKVSCVFFDGCAMLSFESLAETSFARKWPCKYSEPR 412
Db 347 LYVANTVSLIANDYVPDKVSCYISDGCAMLSFESLAETSFARKWPCKFSEPR 406

Qy 413 PEWVFAAKDYLKDVKVSPFVDRRAMKREYEFKIRINALVSKALKCPECSWVMDGTP 472
Db 407 PEWVFAAKDYLKDVKVSPFVDRRAMKREYEFKIRINALVSKALKCPECSWVMDGTP 466

Qy 473 WPCNNTGHPGMQVFLQNGGDAEENELPVLVSYREKRGQFHKKAGMNALVRS 532
Db 467 WPCNNTGHPGMQVFLQNGGDAEENELPVLVSYREKRGQFHKKAGMNALVRS 526

QY 533 AVLNGPFITNLDCOHYINNSKALREAMCFLMDPNLGKQVYQFOPRDGIDKNDRYAN 592
 QY 557 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
 Db 527 AVLTNAPFMUNLDCDHYINNSKAREAMCFLMDPOTGKCYVQFOPRDGIDTHDRYAN 586

QY 593 RNTYFFDINLURGLGIOGYVGGCVFERTALGYEPPIKVKIKKPSLS--KLCGS 649
 Db 587 RNTYFFDINNKGGLDQIOPVYVGTCVFRQALQYGNP--KGPKRKPWVSCDCCPFGS 644

QY 650 RKKNSKAKESKKSGRHDTSTVVFNLDDIEGVEGAFDEKALLMSLEKRFSG 709
 Db 645 RKY-----KERN-----DANGERASLGMDDEKEVILSQMNEKKFEG 683

QY 710 SAVFVASTLMENGVPPSATPENLKEATHIVISCGYEDISDWGMPIGWYGVTEILTIG 769
 Db 684 SSIFVTSTLMEGGVPPSSPAULKEATHIVISCGYEDKTEWGLWLGWLYGSITBDLTG 743

QY 770 FKMHRGWRWISIYCMPKLPKPGSATINLSDRNQVLRWALGSVELLSRICPIANG 828
 Db 744 FKMHCGRGWSIYCMPKRAFKGTAPINLSDRNQVLRWALGSVELLSRICPIANG 803

QY 829 RLKELERFAYNTTYPISIPLUMCTLAUCLETNQFPOISNIASTWLSLSTIF 888
 Db 804 KWKWLREFAVNTTYPISIPLUMCTLAUCLETNQFPOISNIASTWLSLSTIF 863

QY 889 ATGLEMLRNSGVGIDEWMRNEQFWVIGGSALHALFAVFOG3TILKVLAGIDNTF/ESADE 948
 Db 864 ATGIELKNGSYSTEEMWWRNEQFWVIGGSALHALFAVIQGLKLVLAGIDNTF/ESADE 922

QY 949 DGDFAELYFKWTTLIPPTLLTINLVGIVAGSYAINGYQSGWPLGKALEFAFWVITY 1008
 Db 923 DEERGELYTFRKWTTLIPPTLLTINLVGIVAGSYAINGYQSGWPLGKALEFAFWVITY 982

QY 1009 HLYPLKGMSGQRNTPPTIWWSLASSFLSLWWRIPFTSRVGPDLLEGINC 1065
 Db 983 HLYPLKGMSGQRNTPPTIWWSLASSFLSLWWRIPFTSRVGPDLLEGINC 1039

RESULT 9
 US-09-900-237-32
 ; Sequence 32, Application US/09900237
 ; Patent No. US2002120124A1

; GENERAL INFORMATION:
 ; APPLICANT: Allen, Stephen
 ; TITLE OF INVENTION: Plant Cellulose Synthases
 ; FILE REFERENCE: BB1170 US CIP
 ; CURRENT APPLICATION NUMBER: US/09/900, 237
 ; CURRENT FILING DATE: 2001-07-06
 ; PRIOR APPLICATION NUMBER: 60/092, 844
 ; PRIOR FILING DATE: 1998-07-14
 ; PRIOR APPLICATION NUMBER: PCT/US99/15871
 ; PRIOR FILING DATE: 1999-07-13
 ; PRIOR APPLICATION NUMBER: 09/720383
 ; PRIOR FILING DATE: 2000-12-21
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO: 32
 ; LENGTH: 701
 ; TYPE: PRT
 ; ORGANISM: Gossypium hirsutum
 ; US-09-900-237-32

RESULT 9
 US-09-900-237-32
 ; Sequence 32, Application US/09900237
 ; Patent No. US2002120124A1

; GENERAL INFORMATION:
 ; APPLICANT: Stephen
 ; TITLE OF INVENTION: Plant Cellulose Synthases
 ; FILE REFERENCE: BB1170 US CIP
 ; CURRENT APPLICATION NUMBER: US/09/900, 237
 ; CURRENT FILING DATE: 2001-07-06
 ; PRIOR APPLICATION NUMBER: 60/092, 844
 ; PRIOR FILING DATE: 1998-07-14
 ; PRIOR APPLICATION NUMBER: PCT/US99/15871
 ; PRIOR FILING DATE: 1999-07-13
 ; PRIOR APPLICATION NUMBER: 09/720383
 ; PRIOR FILING DATE: 2000-12-21
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO: 32
 ; LENGTH: 701
 ; TYPE: PRT
 ; ORGANISM: Gossypium hirsutum
 ; US-09-900-237-32

RESULT 10
 US-09-838-539-6
 ; Sequence 6, Application US/09838539
 ; Patent No. US2002129401A1

; GENERAL INFORMATION:
 ; APPLICANT: Stalker, D. et al.
 ; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
 ; TITLE OF INVENTION: Sequences
 ; FILE REFERENCE: 15621/03/JS
 ; CURRENT APPLICATION NUMBER: US/09/838, 539
 ; CURRENT FILING DATE: 2001-04-18
 ; PRIOR APPLICATION NUMBER: 60/029, 987
 ; PRIOR FILING DATE: 1996-10-29
 ; PRIOR APPLICATION NUMBER: 08/960, 048
 ; PRIOR FILING DATE: 1997-10-29
 ; NUMBER OF SEQ ID NOS: 12
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO: 6
 ; LENGTH: 974
 ; TYPE: PRT
 ; ORGANISM: Gossypium hirsutum
 ; US-09-838-539-6

Query Match 61.4%; Score 3483.5; DB 10; Length 701;
 Best Local Similarity 90.9%; Pred. No. 3.5e-291; Mismatches 637; Conservative 42; Indels 1; Gaps 1;

QY 366 DYPYDKVSYCFVFDGAGMASFESIETSFARKWPKCKYSISPRAPMYFAAKDYLK 425
 Db 1 DYPYDKVSYCFVFDGAGMASFESIETSFARKWPKCKYSISPRAPMYFAAKDYLK 60

QY 426 DKVQPSFKDRRAMKREYEFKIRKINALSKALKCPEEJWMQDCTPWCONNTGDHPGM 485
 Db 61 DKVQPSFKDRRAMKREYEFKIRKINALSKALKCPEEJWMQDCTPWCONNTGDHPGM 120

QY 486 QVFLGONGGLDAEGBLPLRVYVSBEKRGPFQHKKAGAMNALVRSASVLTNGPTINLD 545
 Db 121 QVFLGONGGLDAEGBLPLRVYVSBEKRGPFQHKKAGAMNALVRSASVLTNGPTINLD 180

QY 546 CDHYINNSKALREAMCFLMPNGLGKQVYQFOPRDGIDKNDRYANRNTYFEDINRLG 180
 Db 181 CDHYINNSKALREAMCFLMPNGLGKQVYQFOPRDGIDKNDRYANRNTYFEDINRLG 240

QY 606 DGIQSPVYVGTCVNRITALGYEPPIKVKHKPSLISKLOGGSKNSK-AKKEDKKK 664
 Db 241 DGIQSPVYVGTCVNRITALGYEPPIKVKHKPSLISKLOGGSKNSK-AKKEDKKK 300

QY 665 SGRHTOSTVPVNLDIEGVEGAGSDEKALLMSLEKREGOSAVFVASTLMENGV 724
 Db 301 SGKHVDSTIVPVLNLDIEGVEGAGSDEKALLMSLEKREGOSAVFVASTLMENGV 360

QY 725 PPSATPENLKEATHIVISCGYEDKDWGMPIGWYGVTEILTIGKPMHARGWRSTYCM 784
 Db 361 POSATPENLKEATHIVISCGYEDKDWGMPIGWYGVTEILTIGKPMHARGWRSTYCM 420

QY 785 KLPFKGSAPINLSDRNQVLRWALGSVELFSRSCPIWYNGKLFLERPAYNTY 844
 Db 421 KRPKGSAPINLSDRNQVLRWALGSVELFSRSCPIWYNGKLFLERPAYNTY 863

QY 845 PITSPILUMCTULLACVLFTNQFPOISNIASTWLSLSTIFATGILEMRWVGIDE 904
 Db 481 PVTAIPLLMCTPLPVALCLTNKFIPOISNIASTWLSLSTIFATGILKMKWNGVQD 540

QY 905 KWRNEOFWVIGGSALHALFAVFOGLKLVLAGIDNTF/VTSKASDEODFABLYMFWTLL 964
 Db .541 KWRNEOFWVIGGSALHALFAVFOGLKLVLAGIDNTF/VTSKASDEODFABLYMFWTLL 600

QY 965 IPPTLLIVNLGVWAGVSTAINSSYQSGWPLGKLFARFWVIVLYPELKGLMRQNT 1024
 Db 601 IPPTLLIVNLGVWAGVSTAINSSYQSGWPLGKLFARFWVIVLYPELKGLMRQNT 660

QY 1025 PTIVVVWSVILLASLWWRIDPITSRVGPDLLEGINC 1065
 Db 661 PTIVVVWSVILLASLWWRIDPITSRVGPDLLEGINC 701

QY 72 LKGSPATIPGDKDGLADEGTVEFNYPOKEK1SERMGWHLTRGKGEMGEPOYDKEVSH 131 ; GENERAL INFORMATION:
 ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| ; APPLICANT: Allen, Stephen
 Db 55 -GSP---YDENLDD-----VERAKGQDSTMMAHL--NKSQDG-----IHA 90 ; TITLE OF INVENTION: Plant Cellulose Synthases
 QY 132 NHPLRPLTSRQDSEGESASASPERLSVSSTAGKRPLYSDVNQSPRRIVPVGNGVA 191 ; FILE REFERENCE: BB1170 US CIP
 QY ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| ; CURRENT APPLICATION NUMBER: US/09/900,237
 Db 91 RHSSVSTDSEADN-----GNS 111 ; CURRENT FILING DATE: 2001-07-06
 QY 192 WKERUDGKMKMKGKNTGPVSTOAASERGGVDIDASTDILADEALLND--EAROPUSRK 247 ; PRIORITY NUMBER: 60/092,844
 Db 112 WKNRVESEPKENKKPAPTKVERAE-----IPPEQOMEKPPAPASQPLSTI 161 ; PRIORITY FILING DATE: 1998-07-14
 QY 248 VSPSSRINPYRQVIMRLVILCFLAHRTNEVPAEALWVSVICBWFWSWLDQF 307 ; PRIORITY APPLICATION NUMBER: PCT/US99/15871
 Db 162 IPIPKSRLAPIRVIVMLLILGLFFHVRVNPDSAFGLWLTSVICBWFWSWLDQF 221 ; PRIORITY FILING DATE: 1999-07-13
 QY 308 PKWPVPNRETYLDRALRYDRGEGPSOLAVDIFVSTWDPLAEPPLVANTVLSILADVY 367 ; PRIORITY FILING DATE: 2000-12-21
 ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| ; NUMBER OF SEQ ID NOS: 33
 Db 222 PKWPVPNRETYLDRSLARYEREGEDELAVIDFVSTWDPLKEPLLTANTVLSILADVY 281 ; SEQ ID NO: 18
 QY 368 PVDKVSCKVFDGAMLSFESLAETSEPARRKWPFCKKYSIERRAPEWYFAAKIDYKD ; LENGTH: 793
 Db 282 PVDKVSCYTSIDOGAAMLTPEFETVADPARKWPFCKKESIERRAPEWYFSQKIDYKD 341 ; TYPE: PRT
 QY 428 VOFSPVKDRRAMREYEEFKIRINALSKALCPEEGVMQDGTPWGPNTRGHPGMQV 487 ; ORGANISM: Glycine max
 ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| ; US-09-900-237-18
 Db 342 VOFSPVKDRRAMREYEEFKIRINALSKALCPEEGVMQDGTPWGPNTRGHPGMQV 401 ;
 QY 488 FLGONGGLDAEENELPRLPVYVEREKKRGEFOHRHKAGAMNLYRVSAYTINGPILNIDCD 547 ;
 ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| ;
 Db 402 FLGYSGARDIEGNELPRLPVYVEREKKRGEFOHRHKAGAMNLYRVSAYTINGPILNIDCD 461 ;
 QY 548 HYJNNSKALREAMCFMLDPNPGKQCVYQFOPQRFDGKDNDYKANRWWFFINLRGDG 607 ;
 ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| ;
 Db 462 HYVNNSKAVAREAMCFMLDPQVGRDVCVYQFOPRGDIDSRDYZANRNTVFFDNMKGDG 521 ;
 QY 608 IGPVWVKGTCVNRNTALGYERPIKVKKPSLSSKGGSRKNSKAKKEDSKKSGR 667 ;
 ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| ;
 Db 522 IGPVWVKGTCVNRNTALGYERPIKVKKPSLSSKGGSRKNSKAKKEDSKKSGR 679 ;
 QY 668 HTDSTVPUVNLDDIEEGVEGAGFDD-EKALIMSMSLEKREFSQSAVETASTLMEENGVPP 726 ;
 ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| ;
 Db 580 REELDAIAFNRLREID---NDEYERSMSLTSFERTFGLSSVTESTMENGVAE 633 ;
 QY 727 SAPPENLIKAEATHVISCGYEDKSDWKGHEIGWIGSYWEIDLITGFKMARGWWSIYCMLKL 786 ;
 ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| ;
 Db 634 SANPSTLKKAEATHVISCGYEEKTAWGKEIGWIGSYWEIDLITGFKMARGWWSIYCMLR 693 ;
 QY 787 PARKGASPINLSDRNQULRWAQSVELFLSRHCPWYFGGGRLKMQPLAYINTVPP 753 ;
 ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| ;
 Db 694 PARKGASPINLSDRNQULRWAQSVELFLSRHCPWYFGGGRLKMQPLAYINTVPP 753 ;
 QY 846 ITSPILMLCTUAVCLIFTNQFTIPOTSNIASTWFLSFLSTFATGILEMRMSGVGIDEW 905 ;
 ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| ;
 Db 754 ITSPILMLCTUAVCLIFTNQFTIPOTSNIASTWFLSFLSTFATGILEMRMSGVGIDEW 905 ;
 ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| ;
 Db 754 ITSPILMLCTUAVCLIFTNQFTIPOTSNIASTWFLSFLSTFATGILEMRMSGVGIDEW 813 ;
 QY 906 WRNEQFWTGIGYSAHLPAFWFOGTLKVLAGIDNFNTVSKASEDGDFAEFLYFKWTLLI 965 ;
 ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| ;
 Db 814 WRNEQFWTGIGYSAHLPAFWFOGTLKMLGIDNFNTVKAAD-DADFGELYIWKWTLLI 872 ;
 QY 966 RPPTLITVNLGVWAGSYAISNGYQWGLGKFLRAFWVHLVFLKLMGRQRTP 1025 ;
 ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| ;
 Db 873 RPPTLITVNLGVWAGSYAISNGYQWGLGKFLRAFWVHLVFLKLMGRQRTP 932 ;
 QY 1026 TIVVWWSVSLASFLSLWWRIDPFTSRVTPGDLE--CGINC 1065 ;
 ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| ;
 Db 933 TIVVWWSVSLASFLSLWWRIDPFTSRVTPGDLE--CGINC 974 ;

RESULT 11 ;
 US-09-900-237-18 ;
 Sequence 18, Application US/09/900,237 ;
 Patent No. US20020120124A1 ;
 Qy 933 AGDTNTFTVTSASDEGDFAEFLYFKWTLLIUPPTLITVNLGVWAGSYAISNGYQ 992 ;
 ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| ;
 Db 660 AGVNTFTVSKAAD-DGEFSLELYFKWTLLIUPPTLITVNLGVWAGSYAISNGYQ 718 ;
 Qy 993 WSPLEFKLFFAFWVWVHLYPFKLGMQRNKRPTIVVWSVSLASFLSLWWRIDPFTSR 1052 ;

RESULT 12
 US-09-900-237-24
 ; Sequence 24, Application US/09900237
 ; Patent No. US200201212A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Allen, Stephen
 ; TITLE OF INVENTION: Plant Cellulose Synthases
 ; CURRENT APPLICATION NUMBER: US/09/900, 237
 ; CURRENT FILING DATE: 2001-07-06
 ; PRIOR APPLICATION NUMBER: 60/092, 844
 ; PRIOR FILING DATE: 1998-07-14
 ; PRIOR APPLICATION NUMBER: PCT/US99/15871
 ; PRIOR FILING DATE: 1999-07-13
 ; PRIOR APPLICATION NUMBER: 09/720383
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO: 24
 ; LENGTH: 740
 ; TYPE: PRT
 ; ORGANISM: Picramnia pentandra
 US-09-900-237-24

Query Match 52.9%; Score 3003;5; DB 10; Length 740;
 Best Local Similarity 73.5%; Pred. No. 8; 9e-260;
 Matches 547; Conservative 94; Mismatches 90; Indels 13; Gaps 5;

Oy 328 REGEPSQDIAADVIFVSDWDPKREPPYTAINTVSLIADWYDVKVSCYVFDDGAAMSLFE 387
 Db KEKPSELALGDIFVSTPDMPKEPLITANTVSLIADWYDVKVTCYVSDDGAAMLTE 62

Oy 388 SLAETSEFARKAWPKCKYSIERTAPERPEWFAAKIDYDKVOTSFVKDRRAMKREYBEEKF 447
 Db 63 ALSETSEFARKAWPKCKFSIERTAPERPEWFAKSQMDYJKNKVHPHSFVRERAMCF 122

Oy 448 IRNALYNSKALKCPEEGWMQDGTTPWPGNNNTGDBHGMTOVFLGQONGGLDAEGLNPLRVY 507
 Db :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||

Db 538 LLLCKKFIVEISNYASILEMFLIFTAATMSILEMQVGCVGIDDMRWNQFWVIGVSSHL 597
 QY 922 FAYVEQILKVLAGIDNTIVTAAKSDECDGFDAEYLKFRWTLLIPPTLLINLVWVAG 981
 Db 598 FALEOGLKLVLAGVNFTVAADE-GDFSELYLFWTTLLPPTLLINLVWVAG 656
 QY 982 VSYAINGSYOSWGLFGLLFFATWVWVHYPFLKGMCQRNRTPTIVWSVLAISFL 1041
 Db 657 VSDAINNGYDSWGLFGLLFFATWVWVHYPFLKGQDRPTIIWWSILLASILT 716
 QY 1042 LWVRIIDPTESRVGPDIJECGIN 1055
 Db 717 LWVRIIDPTESRVGPDIJECGIN 739

RESULT 13
 US-09-900-237-31
 ; Sequence 31, Application US/09900237
 ; Patent No. US200201212A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Allen, Stephen
 ; TITLE OF INVENTION: Plant Cellulose Synthases
 ; CURRENT APPLICATION NUMBER: US/09/900, 237
 ; CURRENT FILING DATE: 2001-07-06
 ; PRIOR APPLICATION NUMBER: 60/092, 844
 ; PRIOR FILING DATE: 1998-07-14
 ; PRIOR APPLICATION NUMBER: PCT/US99/15871
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO: 31
 ; LENGTH: 685
 ; TYPE: PRT
 ; ORGANISM: Gossypium hirsutum
 US-09-900-237-31

Query Match 49.5%; Score 2808; DB 10; Length 685;
 Best Local Similarity 74.4%; Pred. No. 5; 6e-233;
 Matches 512; Conservative 77; Mismatches 77; Indels 22; Gaps 6;

Oy 397 RKWVFPCKRKYSIERTAPERPEWFAAKIDYDKVOTSFVKDRRAMKREYEKFTRINALVK 456
 Db 61 AOKKPEEGWMQDGTTPWGNNTRDGPMTVYLGASAGLVDDGKELPRUVYSSREKRPY 120
 1 RRWVFPCKRKHVERAPERFYNEKIDYDKVHSFVKERRAMKREYEKFVKNALVK 60

Oy 457 ALCKPEEGWMQDGTTPWGNNTRDGPMTVYLGASAGLVDDGKELPRUVYSSREKRPY 516
 Db 62 1 OHHKAGAMNALYRVSAYLTNGPFLNLDCDHVNNSKALREAMCFMLDNPGLQKVQVQ 576

Oy 577 FPQRFIDKNDRAKANRNTVFDJINLRGLDGTQCPVYVTGCVFNRTALGYBEPTRVKH 636
 Db 181 FPQRFIDKNDRAKANRNTVFDJINMLGLDGLQCPVYVTGCVFNROALYQIDPFVSEKR 240

Oy 121 OHHKAGAENALYRVSAYLTNGPFLNLDCDHVNNSKALREAMCFMLDNPGLQKVQVQ 180

Oy 577 FPQRFIDKNDRAKANRNTVFDJINLRGLDGTQCPVYVTGCVFNRTALGYBEPTRVKH 636
 Db 241 PKWICDCWPWCSCCCGSKRSKKRSGKRRGGKLGGLYGGKKKMGKVNKGSSAVFDL 300

Oy 679 DDTEBEGVEGAGFDD-EKALLMSOMLSKREFGOSAVFVASTLMENGGVPSATPENLKEA 737
 Db 301 EEEFEGLE--GYBELEKSTPLMSQRNFERFGDFPQVFLAESTLMENGGVPSATPENLKEA 358

Oy 738 IHWVSCGYEDKSDMGMEIGWYIYSSVTDITLGTRMARGWRSYCMPLKAFFGKAPGSAPINL 797
 Db 359 IHWVSCGYEDKSDMGMEIGWYIYSSVTDITLGTRMARGWRSYCMPLKAFFGKAPGSAPINL 418

Oy 798 SDRUNQVLRWAGSVEILFSRHCPIWGYNGRLKFLERFAYVNNTIVPITSPLMCTL 857

419 SDRLHQVLRWALGSVEIFLRSRHCPLWYGYGKLIKWLRLAYINTTYPFTSIPLLAYCT 478
Db ||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 858 LAVCLIFTNQFIIQIPIQINSIASIWLSLFLSISATQILEMWRMSGVQIDEWMRNEQFWVGIG 917
Db ||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
479 PAVCLGKFIITPLNTSLNTSWPLAFLSUTLNGVLEURSGVSIQDMWRNEQFWVGIG 538
Db ||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
918 SAHLFAVFOGLKLVLAGIDNTFVSKASDEDDGFAELXLFKWTTLIPPTTLVNLVG 977
Db ||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
539 SAHLFAVFOGLKLVLAGIDNTFVSKASDEDDGFAELXLFKWTTLIPPTTLVNLVG 597
Db ||||:|||||:|||||:|||||:
978 WVAGSYA-NSGKQSWGPLFGKLFEEFAFWVYHPLFLKGMLGRNRTPTIVWMSILLA 1037
Db ||||:|||||:|||||:|||||:|||||:|||||:|||||:
598 WVAGSYDAINNGGSGWGPLFGKLFEEFAFWVYHPLFLKGMLGRNRTPTIVWMSILLA 657
Db ||||:|||||:|||||:
1038 IFSLIWVRIDPFTSRVTPDILEGGINC 1065
Db ||||:|||||:
658 IFSLIWVRIDPFLPKQGPVLUKGVEC 685
Db ||||:|||||:

RESULT 14
US-09-838-539-7
; Sequence 7, Application US/09838539
; Patent No. US20020129401A1
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; FILE REFERENCE: 11621/03/US
; CURRENT APPLICATION NUMBER: US/09/838,539
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/029, 987
; PRIOR FILING DATE: 1996-10-29
; PRIOR APPLICATION NUMBER: 08/960, 048
; PRIOR FILING DATE: 1997-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO: 7
; LENGTH: 685
; TYPE: PRT
; ORGANISM: Gossypium hirsutum
; US-09-838-539-7

Query Match 49.4%; Score 2803; DB 10; Length 685;
Best Local Similarity 74.5%; Pred. No. 1.5e-23; Indels 22; Gaps 6;
Matches 512; Conservative 77; Mismatches 76; Indexes 22; Gaps 6;

Qy 396 ARKKWPPCKKYSIPEPAPWYFAAKIDLKDQVQTSVYKDRRAKMRREVEEPKIRINALVS 455
Db ||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 ARKKWPPCKKHNVERAPEFYNEKIDYLKDVKHPFSYKERRAMKREEEFKVIRINALVA 60
Db ||||:|||||:|||||:|||||:|||||:|||||:
456 KALKCPEPGWVMQDGTPWPMQGNNTGDPGMQVFLQNGQGLDAEGNELPRLVYSREKRPG 515
Db ||||:|||||:|||||:|||||:|||||:|||||:
61 KAQQKPEPGWVMQDGTPWPMQGNNTDHPGMQVFLQSGALDVQKEELPRLVYSREKRPG 120
Db ||||:|||||:|||||:
516 FOHKHKAGANALVRSAVLTQGQPFLNUCDAYINNSKALREACMFLMDPNQGKQCVY 575
Db ||||:|||||:|||||:
121 YOHHKHKAGANALVRSAVLTQGQPFLNUCDHINNSKAMREACMFLMDPQFGKKLCVY 180
Db ||||:|||||:
576 QFPQRFGIDKNDRYANNTVFDINLNGLDGQGPVWVGTGCYFNRFALYGEPPRKV 635
Db ||||:|||||:
181 QFPQRFGIDKNDRYANNTVFDINLNGLDGQGPVWVGTGCYFNRFALYGEPPRKV 240
Db ||||:|||||:
678 LODIEEGEGAGBDD-EKALLMSOMSLKREGOSAVAFSTLMEGGVPPSATPENLKE 736
Db ||||:|||||:
301 LEERIEGLE-GYEELERSTMSOKNFRERFGOSPVFTASTLMEGGLPEGTNSTSLIKE 358
Db ||||:|||||:
737 AHWISCGYEDKSDWGMGEMIGWIWGVSTEDILGFKMHARGWSIYCMPLPKPARKGASPIN 796
Db ||||:|||||:

RESULT 15
US-09-900-237-16
; Sequence 16, Application US/09900237
; Patent No. US20020120124A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen
; TITLE OF INVENTION: Plant Cellulose Synthases
; FILE REFERENCE: BBI170 US CIP
; CURRENT APPLICATION NUMBER: US/09/900, 237
; CURRENT FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 60/092, 844
; PRIOR FILING DATE: 1998-07-14
; PROJ. APPLICATION NUMBER: PCT/US99/15871
; PRIOR FILING DATE: 1999-07-13
; PRIOR APPLICATION NUMBER: 09/720383
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Microsoft Office 97
; SEQ ID NO: 16
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Glycine max
; US-09-900-237-16

Query Match 46.6%; Score 2647.5; DB 10; Length 610;
Best Local Similarity 78.8%; Pred. No. 3.1e-219; Indels 3; Gaps 2;
Matches 482; Conservative 55; Mismatches 72; Indexes 3; Gaps 2;

Qy 455 SKALKCPEEGWVMQDGTPWPMQGNNTGDPGMQVFLQNGQGLDAEGNELPRLVYSREKRPG 515
Db ||||:|||||:|||||:
1 AKAQKMPPEGWTMDQTPWMCNNPDRHGMQVFLHSGGDLTDGNELPVUVYSREKRPG 60
Db ||||:|||||:
515 GFOHKHKAGANALVRSAVLTQGQPFLNUCDHINNSKALREACMFLMDPNQGKQCVY 575
Db ||||:|||||:
61 GFOHKHKAGANALVRSAVLTQGQPFLNUCDHINNSKALREACMFLMDPNQGKQCVY 120
Db ||||:|||||:
575 WQPQRFGIDKNDRYANNTVFDINLNGLDGQGPVWVGTGCYFNRFALYGEPPRKV 634
Db ||||:|||||:
121 WQPQRFGIDKNDRYANNTVFDINLNGLDGQGPVWVGTGCYFNRFALYGEPPRKV 240
Db ||||:|||||:
635 KHKKPSLSSK-CGGSRKNSKA-KRSDSKKSRGHDTSTWPVNLDIEEGVEAGFDDE 693
Db ||||:|||||:
181 EDLERPNIVKSCSGSRKSKKGEKKGLLGYGKKKKMKNVYKKGSSAPVD 238
Db ||||:|||||:
694 KALLMSOMSLKREGOSAVAFSTLMEGGVPPSATPENLKEAHWISCGYEDKSDWGM 753
Db ||||:|||||:
239 RTLMQSOKSLKREGOSPVFTATMEQGIPPSNPATLKEAHWISCGYEDKTEWGK 298
Db ||||:|||||:
754 EIGWIGSVTBDILGFKMHARGWSIYCMPLPKPARKGASPINSDRLQVLRWALGSVE 813
Db ||||:|||||:

Db	299	EIGWIGSVTEDILTGKFMHARGWISIYCMPPRFAPKGASPINLSDRLNQVLRWALGSIE	358
Qy	814	IIFSRHCPIWGYNGRUKFLERFAYVNITIPTSPILMCYCLAVLFTNOFIPOIS	873
Db	359	IFLSRHCPLWGYNGKLUKPLMLRAYINIVYPTSPILAYCTLPACFLTNKFLPEIS	418
Qy	874	NIASIWFLSLSLTSIAFGILEMWGSGVIDEMARNEOFWVIGVS AHLFAVGOLIKVLA	933
Db	419	NFASMWFLVLSISIFTSILELRWSGVSIEDWMARNEOFWVIGTS AHLFAVGOLIKVLA	478
Qy	934	GIDTNFVUTSKASDEDDFAELYFKMWTIJPPTLLIVLNLYVGWGVSYAINSGQSW	993
Db	479	GIDTNFVUTSKASDEDDFAELYFKMWTIJPPTLLIVLNLYVGWGVSYAINSGQSW	538
Qy	994	GPLFGKLFKAFWIVHLYPFLKGLMGRQNRTPTIVVVWSVLASIFSLWVRLDPTSRV	1053
Db	539	GPLFGKLFKAFWIVHLYPFLKGLLGRONRPTIVVVWSVLASIFSLWVRLDPTSDS	598
Qy	1054	TGPDILCGINC 1065	
Db	599	NKLTN5OCCINC 610	

Search completed: June 16, 2003, 11:15:47
Job time : 36 secs

THIS PAGE BLANK (USPTO)

Copyright (c) 1993 - 2003 Compugen Ltd. GenCore version 5.1.6 OM protein - protein search, using sw model

ALIGNMENTS

28	101.5	1.8	765	2	US-08-825-886-19	Sequence 19, Appl
29	101.5	1.8	900	1	US-08-425-061-20	Sequence 20, Appl
30	101.5	1.8	900	2	US-08-825-886-20	Sequence 20, Appl
31	101.5	1.8	914	1	US-08-425-061-21	Sequence 21, Appl
32	101.5	1.8	914	2	US-08-825-886-21	Sequence 21, Appl
33	101.5	1.8	1202	1	US-08-425-061-22	Sequence 22, Appl
34	101.5	1.8	1202	2	US-08-825-886-22	Sequence 22, Appl
35	101.5	1.8	1363	1	US-08-425-061-23	Sequence 23, Appl
36	101.5	1.8	1363	2	US-08-825-886-23	Sequence 23, Appl
37	101.5	1.8	1852	1	US-08-425-061-24	Sequence 24, Appl
38	101.5	1.8	1852	2	US-08-825-886-24	Sequence 24, Appl
39	101.5	1.8	1863	1	US-08-598-591-2	Sequence 2, Appl
40	101.5	1.8	1863	1	US-08-598-591-2	Sequence 2, Appl
41	101.5	1.8	1863	1	US-08-483-553-2	Sequence 2, Appl
42	101.5	1.8	1863	1	US-08-487-002-2	Sequence 2, Appl
43	101.5	1.8	1863	1	US-08-483-554B-2	Sequence 2, Appl
44	101.5	1.8	1863	1	US-08-798-691-2	Sequence 2, Appl
45	101.5	1.8	1863	1	US-08-798-691-4	Sequence 4, Appl

PRIOR FILING DATE: 1996-10-29
 NUMBER OF SEQ ID NOS: 12
 SOFTWARE: FastSEQ for Windows Version 4.0
 SEQ ID NO 7
 LENGTH: 685
 TYPE: PRT
 ORGANISM: Gossypium hirsutum
 US-08-960-048-7

Query Match 49 4%; Score 2803; DB 4; Length 685;
 Best Local Similarity 74.5%; Pred. No. 5.9e-276; Mismatches 76; Indels 22; Gaps 6;
 Matches 512; Conservative 77; MisMatches 76; Indels 22; Gaps 6;

Qy 396 ARKKWPFCKKVSIERAPPEFAAKIDYLKDKVQNSFKVKDRAMREYEEFKIRINALVA 455
 Db 1 ARRWPFPCKKHNRVERAPPEFVNKEIDYLKDKVHFSFKERRAMREYEEFKIRINALVA 60

Qy 456 KALKCPEEGWMQDGTPWPGNNTDHPGMIVOLYIGSAGALWDGKELPLRVYVSREKRP 515
 Db 61 KAQQKPEEGWMQDGTPWPGNNTDHPGMIVOLYIGSAGALWDGKELPLRVYVSREKRP 515

Qy 516 FOHHKAGAMALVRYSAVLTINGPFLNLQDHYINNSKALREAMCFLMOPNLQKVOCVV 575
 Db 121 YOHHKAGAENALVRYSAVLTINGPFLNLQDHYINNSKANREAMCFLMOPNLQKVOCVV 180

Qy 576 QPPQRDGIDKNDRKYANRNTVFFDINLRGLDGIQGVYVGTCVFNRTALGYEPPIKV 635
 Db 181 QFPQREQDIDRHDRYANRNVVFDDINMLGLDGLQGVYVGTCVFNRLQYDPYSEK 240

Qy 636 HKK-----PSILSKLUGGSRSRKNSKAKE-----SDKKSGRH--TDSTVPVFN 677
 Db 241 RPKMTQDCWPSWCWCCCCGGSRKSKRGKKGGLGGLYGKKKKMKGKVVKGSAVPFD 300

Qy 678 LDDIEGEGAGFDD-EKALLMSQSLEKRGQSAVFSPLMENGVPSATPENLKE 736
 Db 301 LEEIEEGLE--GYEELEKTSMSQNKFRKGQSPVIASTLMEENGVLPCBTNSTLSIKE 358

Qy 737 AIHVSCGYEDKSDMOWMEIGIYGSYTEDILTGFKHARGRSITCAMPKLAFAKGASPIN 796
 Db 359 AIHVSCGYEKTEMGKEIWIYGVSYTEDILTGFKHARGRSVCPKRAFKGASPIN 418

Qy 797 LSDRQLNQVLRALGSVEILSRSRHCPIWYGYNGRLKFLERAYVNTIYTPTISIPLMYCT 856
 Db 419 LSDRQLHQVLRMALGSVEILSRSRHCPIWYGYGGKLWLERAYINTIVPFISIPLAYCT 478

Qy 857 LLAVCIFTNOTIPOSNASIWIUSLFLSIFATGLEMRMSGVGDEDEWRNEQFWIGG 916
 Db 479 IPIAVCLTGKIIPTISPLNTSWSWFLAFLLSTIATGYLTLRMSGVSIODWRNEQFWIGG 538

Qy 917 VSAHLEAVFOGLKVLVLAGIDNTFT~~TS~~SKASDEDDGDAEYLFWKWTLLIPPTTLLVNLV 976
 Db 539 VSAHLEAVFOGLKVLVLAGIDNTFT~~TS~~KAAD-DTEFGELVLFKWTTLLIPPTTLLVNLV 597

Qy 977 GVVAGVSYAINSGYGSWGP~~TS~~KAFLPAFWVIVHLVPLFLKGMLGRONTPTIWWVNSVLLA 1036
 Db 598 GVVAGVSYADSAINGYGSWGP~~TS~~KAFLPAFWVILHLYPLFLKGMLGRONTPTIWWVNSVLLA 657

Qy 1037 SIFSLLWVRIDPFTSVTGPIDLEGI 1063
 Db 658 SIFSLLWVRIDPFLPQTGPVLUKQCVV 684

PRIOR APPLICATION NUMBER: 60/029, 987
 PRIOR FILING DATE: 1996-10-29
 NUMBER OF SEQ ID NOS: 12
 SOFTWARE: FastSEQ for Windows Version 4.0
 SEQ ID NO 11
 LENGTH: 693
 TYPE: PRT
 ORGANISM: Escherichia coli
 US-08-960-048-11

Query Match 4 9%; Score 278; DB 4; Length 693;
 Best Local Similarity 19.3%; Pred. No. 7.7e-9; Mismatches 191; Indels 362; Gaps 30;
 Matches 153; Conservative 86; MisMatches 191; Indels 362; Gaps 30;

Qy 260 MYIMMRUVICLICFLHYRIT-----APVNPNAFLMIVSVCIEIWPAWSWIDQPFKWFV 313
 Db 21 MJIVLSLJWVSCRYIWWRTSTLNWDDPVSLVCGLILLFAITYAWTVL-VLGYFOVWPL 78

Qy 314 NRETYLDRLALRYRECEPSQLAQAVDIFVSTVDPLKEPPVLTANTVLSILAVDYPDKVS 373
 Db 79 NRQP---VPIJKMSLWPS---VDIFVPTN---EDLNWVKNTIVASLGIDWKDQLN 127

Qy 374 CVFDDGAAMLSFESLAETSEFARKWVPECKYSIESRAPEWYAAKIDYLKDKVQTSFV 433
 Db 128 IWILDDGG-----
 Db 434 KDRRAMREYEFKIRINALVSKALCPEEGWMQDGTPWPGNNTGDHPGMLQVFLGQNG 493
 Db 136 -----REFROF-----AQNV 146

Qy 494 GLDAEGLNPLRVYVSREKRFQFHKKAGAMNALVRVSAVLTINGPFLNLQDHYINNS 553
 Db 147 GVK-----YIATUT---HEAKAGNTINAJKA-----KGEFVSIIDCDHVPTRS 188

Qy 554 KALREAMCFLMDPNLGKQVYCQVQFQPRF--DGIDKN---DRYANRNTVFFDINLRGLD 606
 Db 189 FIQMTKOWFELKE---KOLAMMQTPHFFSPDPFERNLGRFRKTTPNEGTLFGLVQDND 244

Qy 607 GIQGPVYVGTVGCFVNRTALGYEPPIVKHKRPSLISKCGSRKNSKAKESDRKKSG 666
 Db 245 MNDAFFCGSCAVIRR-----KPI-----
 Db 263 -----LDEI-----
 Db 270 -----AVE-----
 Db 727 SATPENILKEATHVSCGYEDKSDMWSMEIGIYTGSYTEDILTGFKHARGRSVCPK 786
 Qy 787 PAFKGSAPINLSDRNLNOVLRWALGSVEILSRSRHCPIWYGYNGR-LKFLERAYVNTIYT 845
 Db 297 POAAGLATESLSAHIGORIRWARGMVQI-FRDNLPL---TGKGLKAQORLQCYVNAFMHE 351

Qy 846 ITSPILMYCT-----LLAVCIFTNQFIPQI-----SNIASIWFLSFLS 886
 Db 352 LSGIPRULFLTAPLAFLLHAVIYIAPALMTFLPHMITHASLTNSKICKYRISFE 411

Qy 887 IFAT-----GLEMRWMSGVGDEDEWRNEQF--WWIGVSAHFAVE 925
 Db 412 IYETVLAWYIAPTLVALINPHKGKFNTAKGGVVE--EEYVWVVIS--RPIYFLV 465

Qy 926 QGILKVLVLAGIDNTFT~~TS~~SKASDEDDGDAEYLFWKWTLLIPPTTLLV-----NLV 976
 Db 466 LNLVGVAVGI-----
 Qy 977 GVVAGVSYAINS 988
 Db 502 VLGVAVSVES 513

RESULT 5
 US-08-960-048-12
 Sequence 11, Application US/08960048C
 Patent No. 6271443

GENERAL INFORMATION:
 APPLICANT: Staker, D. et al.
 TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
 TITLE OF INVENTION: Sequences
 FILE REFERENCE: 15621/01/US
 CURRENT APPLICATION NUMBER: US/08/960,048C
 CURRENT FILING DATE: 1997-10-29

RESULT 5
 US-08-960-048-12

ATTORNEY/AGENT INFORMATION:
NAME: Murphy, Lisabeth Felix
REGISTRATION NUMBER: 31547

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 393-2286
TELEX: 340817 MACPAG SFO

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3031 amino acids

TYPE: AMINO ACID

TOPOLOGY: Linear

MOLECULE TYPE: Protein

US-07-689-008-2

Query Match 4.0%; Score 228.5; DB 1; Length 3031;
Best Local Similarity 19.3%; Pred. No. 1.2e-12; Mismatches 239; Indels 329; Gaps 30;
Matches 154; Conservative 84; MisMatches 239; Indels 329; Gaps 30;

Qy 263 MRLVILCFLHYRITNPVNAFALW-----VSVICEWFALSWILDQFPKWPVNRE 316
Db 80 VLSALVSLSYLTLWRLTETD-FDTWIGGLGTYLMLAELYALMFLSYEQNTOPLHRR 137

Qy 317 TYDLRLALKYDRGEPSOLAAVDIFVSTWDPLKEPPLVTANTVSLILAWDYPDVKSYCV 376
Db 138 P----LPEDNVDWMP----VDFIPFYD-EQLSIVRLTVLGALGDWPPDKVNVYI 186

Qy 377 FDDGAAMLSFESLAETSEFARKWVPRCKKYSIEPPRWYFAAKIDYLKDQVTSFWKDR 436
Db 187 LDGG----- 190

Qy 437 RAMKREYEEFKIRINALVYKALKCPECGWVMQDGTPWPQGNNTGDPGMQVFLQNGLD 496
Db 191 -VRPFEBQFAKOCGAL----- 205

Qy 497 AEGNELPPLVWVSEKRGFQHHKAGAMNALVRSAVLTINGPFLINIDCDHVINSKAL 556
Db 206 -----YIGRVD---SSHAKAGNHLAIKA-----RTSGDYLIDCDH-IPTRAFL 246

Qy 557 REAMCFMLMDPNLKGKQVYQFOPRFDGIDKNDYANRNTVFFDINLRGLDGHQGPVYGT 616
Db 247 QIANGWWV--ADRKLALMOTPHFYSPDP----- 273

Qy 617 GCVFNRTALGYEPPIKVKKKPSLLSKLCGGSRKKNSKAKKESDKKKSGRHIDSTVPVF 676
Db 274 --FORNLAVGYRTP-----PEG 288

Qy 677 NLDDIEEVSEGAGFDEKKALLMSQMSLERKRGOSAVFASLTMENGGVPPSATPENLKE 736
Db 289 NL---EVGVIQDND-----FMDATEFCGSCAI-----LRR 317

Qy 737 AIHWISCGYEDKSDWGMETGWIYGSVTEIDLTKFMHARGWRWSITYCMLPKPAFKGAPIN 796
Db 318 AIESIG-----GFAVETVDEDAHTALMRQRGWSTAYL-RIPVAGLATER 362

Qy 797 LSDRLNQYERWLALGSVETFSRHCPIWYGNRKLFRFAYVNTTIVPITSPLPLMYCT 856
Db 363 LTHHIGQMRWARGMIOI-FRVNDPM--LGGGLKLKGRLCYLSAMTSFFAIPRVLFLA 418

Qy 857 LLAVCLFNNQFTIPOQISIASTIWLSELSIFATGILERWWS--GIVGIDEMWRNEQEW- 912
Db 419 SPLAFLFFGCONITAASP-----LAVLAYAIPHMHSATAAKWKNGWR-YSFWS 466

Qy 913 VIGGSVAHLFAVQIGLKLWLAGIDTNFTT⁽¹⁾SKA--SDEDGDAELYLIFKWTTLLIBPTT 969
Db 467 EYVETTMALFLVRYTITLMPSKGKENTWKGQVLEEEBFIDGATY-----PNI 516

Qy 970 LLIVNLVGVWGVSYAINGYOSWGPLG-----KLFEEFWIWLYPELKGMGRQNRPPTI 1027
Db 517 IF-----AGIMWTLGLANGFELTEHF-----NOLAGIAKRAYL 550

Qy 1028 VVWWS-----VILASI 1038

Db 551 NCIWAMISIILUAAI 566

RESULT 9
US-08-960-048-9
Sequence 9; Application US/08960048C

Patent NO. 6271443
GENERAL INFORMATION:
APPLICANT: Stalker, D. et al.

TITLE OF INVENTION: Plant Cellulose Synthase and Promoter

TITLE OF INVENTION: Sequences

FILE REFERENCE: 15621/01/US

CURRENT APPLICATION NUMBER: US/08/960.048C

CURRENT FILING DATE: 1997-10-29

PRIOR APPLICATION NUMBER: 60/029, 987

NUMBER OF SEQ ID NOS: 12

SOFTWARE: FastSEQ for Windows Version 4.0

SEQ ID NO 9
LENGTH: 723

TYPE: PRT

ORGANISM: Acetobacter xylinum

US-08-960-048-9

Query Match 3.5%; Score 195; DB 4; Length 723;
Best Local Similarity 20.1%; Pred. No. 1.9e-10; Mismatches 241; Indels 306; Gaps 36;
Matches 161; Conservative 93; MisMatches 241; Indels 306; Gaps 36;

Qy 251 PSSRINPVRMNTMRLVILCFLHYRITNPVNAFALW-----VSVICEWFALSWIL 304
Db 70 PSRSQIE-LEVLSGLVS-L-RYLTWRLHETL-SFDWVLOGLIGTMLVIAEYALMMLFL 125

Qy 305 DQTPKWPVNRETYLDRIALRYDREGEPSQLAAVDIFVSTVPLKEPPLVTANTVSLILA 364
Db 126 SYQTIAFLHRAP---IPL---PPNDEMWPVDFVPTV---EBSLIVRLTVLGSLG 174

Qy 365 VDPVDPKVSCYVDDGAMLSFESLAETSEFARKWVPRCKKYSIEPPRWYFAAKIDYL 424
Db 175 IDWPEPKVRVHILDDG----- 197

Qy 425 KDKVQTSVVKDRRAMKREYEFKIRINALVYKALKCBEQGWTMQDGTPWPGNNTGDIHGM 484
Db 198 -----FAEC----- 202

Qy 485 IQVFLQONGGLDAGENELPRLVYVSREKRGFQOHKKAGAMALVRSAVLTINGPFLINL 544
Db 203 -----GAN-----YIARP-----NEHAKAGNLYAIG---HNGDYTLIFL 235

Qy 545 DCBVHYINNSKALBACMCILM-DPNLKGKOVCYQFQPQRDGTDOKNDYANRNTVFFDINLR 603
Db 236 DCBH-VPTRAFLQTMGMWNVEDP---KIALMOTPFFHFSYSPD----- 273

Qy 604 GLGDIQGPVYVGRCVFIRALGYEPPIKVKKPSLLSKLCGGSRKKNSKAKKESDKK 663
Db 274 -----FORNLSAGYRTP----- 285

Qy 664 KSGRHTDSTVPVNLDIDIEEGVEGAFDEKKALLMSQMSLERKRGOSAVFASLTMENG 723
Db 286 -----RECNL-----FYGWVQDGN-----FWDATFFCGSCAI----- 313

Qy 724 VPPSATPENLKAIAHVISCGYBDKSDWGMETGWIYGSVTEIDLTKFMHARGWRSYCM 783
Db 314 -----LRRTAIEQIG-----GFATQTVTEDAHTALKMORLGWSTAYL- 350

Qy 784 PKLPAFKGKSAPINLSDRINQVLRWALGSVETFSRHSRPIWYGNR-LKFELRFAYNTT 842

Db 351 -RPLAGGLATERLITLHGQYRWRASMQI-FRDINPLF---GRGLSWGORLCYLSAM 404

Qy 843 IPITSIPLMCTILAVLCLFTNOIFIQISIASTIWLSELSIFATGILERWWSVG----- 901

Db 405 TSFLFMPVRFVLISSPLAFLFFGQNTIASP-----LALVAYAIPHM-FHAGV----- 452

Qy 902. ---IDEAWRNEOFW-VIGGSVAHLFAVFOGILKVLLAGIDTNFTW⁽¹⁾KASDEGDFAFYL 957

RESULT 10
 PCT-US91-01726-4
 Sequence 4, Application PC/US9101726
 GENERAL INFORMATION:
 APPLICANT: Brown, Malcolm R
 APPLICANT: Saxena, Indra M
 TITLE OF INVENTION: Recombinant Cellulose Synthase
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: David L. Parker
 STREET: 600 Congress Avenue Suite 2300
 CITY: Austin
 STATE: Texas
 COUNTRY: United States
 ZIP: 78701
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PC/US91/01726
 FILING DATE: 19910314
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 494093
 FILING DATE: 15-MAR-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Parker Mr, David L
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 512-320-7200
 TELEFAX: 512-474-7577
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 723 amino acids
 TYPE: AMINO ACID
 STRANDEDNESS: single
 TOPOLOGY: linear
 PCT-US91-01726-4

Query Match 3.5%; Score 196; DB 5; Length 723;
 Best Local Similarity 20.1%; Pred. No. 1.9e-10; Matches 161; Conservative 93; Mismatches 241; Indels 306; Gaps 36;

Qy 251 PSSRNPYRMVIMRLUVLICLFLHYRINTPNPNAFLW-----VSVICEWFALSWIL 304
 Db 70 PSRRSOFF-LETVLSGLNVL-RYLTWRIETL--SFDIWWLQGLGTMVIAELYALMFL 125

Qy 305 DOPPKWTPPVNRETYLDRLRALDREGEPSQLAANDFVSTVPLKERPLVANTVLSILA 364
 Db 126 SYFQTIAPLHRAP----LPL---PRPDEPTVDFVPTN--ERLSIVLTVLGSIG 174

Qy 365 VDPVPDKVSCYYFDGGAMLSFSELASLTSEFARKWYPPFKKYSIEPRAPEWFAAKIDYL 424
 Db 175 IDWPPEKRVHLDDG-----RRPE--FAA---- 197

Qy 425 KOKVQSFVKDRAMKKEEYKIRINALVSKALKCPEGWMQDGTPWPGNITGHPGM 484
 Db 198 -----FAAC---- 202

RESULT 11
 US-08-119-773-4
 Sequence 4, Application US/08119773
 Patent No. 5460942
 GENERAL INFORMATION:
 APPLICANT: Chou, Janice Y.
 APPLICANT: Lei, Ke-Jian
 APPLICANT: Shelly, Leslie L.
 TITLE OF INVENTION: GLUCOSE-6 PHOSPHATASE: THE GENE AND PROTEIN AND RELATED MUTATIONS
 NUMBER OF SEQUENCES: 36
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend Khourie and Crew
 STREET: Steuart Street Tower, One Market Plaza
 CITY: San Francisco
 STATE: California
 COUNTRY: US
 ZIP: 94105-1493
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/119, 773
 FILING DATE: 10-SEP-1993
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Weier, Kenneth A.
 REGISTRATION NUMBER: 31.677
 REFERENCE DOCKET NUMBER: 15280-175
 TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 543-9600
 TELEFAX: (415) 543-5043
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 357 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ORIGINAL SOURCE: Homo sapiens
 FEATURE: Modified-site
 NAME/KEY: Modified-site
 LOCATION: 1..357
 OTHER INFORMATION: /label= R to C at 83
 OTHER INFORMATION: /note= "The G-6-Pase amino acid sequence mutated
 OTHER INFORMATION: from Arg to Cys at position 83."
 US-08-119-773-4

Query Match 2.2%; Score 124; DB 1; Length 357;
 Best Local Similarity 19.8%; Pred. No. 0 0011;
 Matches 79; Conservative 44; Mismatches 122; Indels 154; Gaps 16;
 QY 736 EATHVISCYEGEKSDFWMEIGWYGSTEDILTGKHHARGRSIYCMPLKPAFKGSAPI 795
 DB 14 QSTHYLQVNQYQDSQDW---FILVSIVADLRNAFYVLFPIW---FHLQEAVGIKLLWA 65
 QY 796 NLSDRNLQVLRAALGSEVILSRHCPWYG----- 825
 DB 66 VIGDWNLNVFKW-----IFGQ-CPYWVWLDTDYNTSVPLIKOPPVTCETGPSSPG 118
 QY 826 -----YNGRLKFLERFAVNTTYPITSPILLMCTTLAVCL--- 862
 DB 119 HAMTAGVYYVWMTSISIFQKIKPYRFRCLNVLW-----LGFWAQVLNCISRY 172
 QY 863 ---FTNOQTIPQINIA-----SIWFSL-----FLSFATGILEMRWSGVG 901
 DB 173 LAAHFPHOVWAGVLSGIATVETFSHHSIYNASLKKYFLITFLFSFAIGFY-LLIKGLG 231
 QY 902 IDEWWRNE-QFWVIGGVSAAHFIAVFGQGILKVLAGIDTNFTVSKASDEGDFAEYLFK 959
 DB 232 VDLIWTLERAKQRMCEOPEWH-----IDMT----- 256
 QY 960 WTLLIIPPTLILVNLGVWAGVSAYAINSG-----YQSWGPFEGKLFFAFWIVHL 1010
 DB 257 -----PFASLKLNL-GTLEGLGALNSSMYRESCKGKLSKWLPCLSSIVASLVLLHV 308
 QY 1011 YPFLKGJLMGRQNPPTIVVWWSLASSIFSLAFLWVRIOPF 1049
 DB 309 FDSLK---PPSQVELVYVFLSFCKSAVPLASVSVIY 343

RESULT 12
 US-08-119-773-6
 Sequence 6, Application US/08119773
 ; Patent No. 5460942
 ; GENERAL INFORMATION:
 ; APPLICANT: Chou, Janice Y.
 ; APPLICANT: Lei, Ke-Jian
 ; APPLICANT: Shelly, Leslie L.
 ; TITLE OF INVENTION: GLUCOSE-6-PHOSPHATASE: THE GENE AND
 ; TITLE OF INVENTION: PROTEIN AND RELATED MUTATIONS
 ; NUMBER OF SEQUENCES: 36
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Townsend and Townsend Khourie and Crew
 ; STREET: Stewart Street Tower, One Market Plaza
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: US
 ; ZIP: 94105-1493
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk

Query Match 2.0%; Score 116; DB 1; Length 357;
 Best Local Similarity 19.5%; Pred. No. 0 0073;
 Matches 78; Conservative 44; Mismatches 123; Indels 154; Gaps 16;
 US-08-119-773-6

Query Match 2.0%; Score 116; DB 1; Length 357;
 Best Local Similarity 19.5%; Pred. No. 0 0073;
 Matches 78; Conservative 44; Mismatches 123; Indels 154; Gaps 16;
 QY 736 EATHVISCYEGEKSDFWMEIGWYGSTEDILTGKHHARGRSIYCMPLKPAFKGSAPI 795
 DB 14 QSTHYLQVNQYQDSQDW---FILVSIVADLRNAFYVLFPIW---FHLQEAVGIKLLWA 65
 QY 796 NLSDRNLQVLRAALGSEVILSRHCPWYG----- 825
 DB 66 VIGDWNLNVFKW-----IFGQ-CPYWVWLDTDYNTSVPLIKOPPVTCETGPSSPG 118
 QY 826 -----YNGRLKFLERFAVNTTYPITSPILLMCTTLAVCL--- 862
 DB 119 HAMTAGVYYVWMTSISIFQKIKPYRFRCLNVLW-----LGFWAQVLNCISRY 172
 QY 863 ---FTNOQTIPQINIA-----SIWFSL-----FLSFATGILEMRWSGVG 901
 DB 173 LAAHFPHOVWAGVLSGIATVETFSHHSIYNASLKKYFLITFLFSFAIGFY-LLIKGLG 231
 QY 902 IDEWWRNE-QFWVIGGVSAAHFIAVFGQGILKVLAGIDTNFTVSKASDEGDFAEYLFK 959
 DB 232 VDLIWTLERAKQRMCEOPEWH-----IDMT----- 256
 QY 960 WTLLIIPPTLILVNLGVWAGVSAYAINSG-----YQSWGPFEGKLFFAFWIVHL 1010
 DB 257 -----PFASLKLNL-GTLEGLGALNSSMYRESCKGKLSKWLPCLSSIVASLVLLHV 308
 QY 1011 YPFLKGJLMGRQNPPTIVVWWSLASSIFSLAFLWVRIOPF 1049
 DB 309 FDSLK---PPSQVELVYVFLSFCKSAVPLASVSVIY 343

RESULT 13
 US-08-119-773-5
 Sequence 5, Application US/08119773
 ; Patent No. 5460942
 ; GENERAL INFORMATION:
 ; APPLICANT: Chou, Janice Y.
 ; APPLICANT: Lei, Ke-Jian
 ; APPLICANT: Shelly, Leslie L.

```

RESULT 14
US-08-119-773-2
; sequence 2, Application US/08119773
; Patent No. 5,609,422
GENERAL INFORMATION:
APPLICANT: Chou, Janice Y.
APPLICANT: Lei, Ke-Jian
APPLICANT: Shelly, Leslie L.
TITLE OF INVENTION: GLUCOSE-6-PHOSPHATASE: THE GENE AND PROTEIN AND RELATED MUTATIONS
TITLE OF INVENTION: PROTEIN AND RELATED MUTATIONS
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Khourie and Crew
STREET: Stewart Street Tower, One Market Plaza
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94105-1493
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/119,773
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15-880-175
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 357 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
; US-08-119-773-2

Query Match.
Best Local Similarity: 2.0% Score: 115; DB 1; Length: 195; Pred. No: 0.0092; Mismatches: 123; Indels: 78; Conservative: 44; Mismatches: 123; Indels: 78; Matches: 78; Conservative: 44; Mismatches: 123; Indels: 78;
Qy 736 EAIIHWISCGYEKDSDMGMEIGWIGVSWEDILTGFKMHARGWRRIYV
Db 14 QSTHYLQLQVNQDQSOW-----FLIVSVIADRNRNFYVLFPIW----F
Qy 796 NLSDRLNQVLRLGVALLGSVEILFSRHCPIWG-----
Db 66 VIGDWLNVLVFKN-----ILFGQR-PYWWVLDWDYYSNTSVPLIKQQ
Qy 826 -----YNGRLKFELRFAYVTTTIVTISPIILL
Db 119 HAMGTAGVYYVMASTSLISQGKIKPTYFRCLNVILW-----LG
Qy 863 -----FNQFILPQISNIA-----SIWFLS-----FISIWI
Db 173 LAAHFPHQHVAGVGLSGIAVTEIFSHINSIYNAKKYFLTEFLFSS
Qy 902 IDEWWRNE-QFWVIGVSAHLEAVFOGLKVLAGIDTNFTVTSKAKA
Db 232 VDLMLLEKAQRMCEQFBWVH-----IDTF-----
Qy 960 WITLILPPTTLIVLNUGVWAGSYAING-----YOSWQRPUL

```

357;
 s 154; Gaps 16;
 CMPKLPAPFKGSAPI 795
 : : |
 HIQEBAVGIKLILVA 65
 ----- 825
 RFPVTCERGPSPSG 118
 MYCITLLAVCLL---- 862
 : | |||
 FWAVQVNQVCLSRIV 172
 FATGILEMRWSQVG 901
 || | : |:
 FAIGFY-LLLKGIG 231
 SDEDGDAELYLFK 959
 ----- 256
 EGKLFEEFWVIVHL 1010

RESULT 15
US-09-000-094-46
QY Sequence 46, Application US/09000094
Db Patent No. 6365160
GENERAL INFORMATION:
APPLICANT: WEBB, Elizabeth Ann
MARGETTS, Mary Brigid
COX, John Cooper
FRAZER, Ian
MCMLIAN, Nigel Alan John
WILLIAMS, Mark Phillip
MOONEY, Margaret Bridget
HOLLAND
EDWARDS, Stirling John Polyprotein Constructs
TITLE OF INVENTION: PAVILLONAVIRUS CONSTRUCTS
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/0000, 094
FILING DATE: 21-Apr-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/AU96/00473
FILING DATE: 26-JUL-1996
APPLICATION NUMBER: AU PN 4439/95
FILING DATE: 27-JUL-1995
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29-768
REFERENCE/DOCKET NUMBER: 017227/0137
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 157 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-000-094-46

Query Match 1.9%; Score 110; DB 4; Length 157;
Best Local Similarity 19.3%; Pred. No. 0; DB 4;
Matches 114; Conservative 77; Mismatches 221; Indels 180; Gaps 30;

QY 2 ESEGETACKPMK--NTVFPOTCICSDNKGTVGDREFAACDICSPVCPCEYERKD 58
Db 877 EVQGDS-QPLKHFQIVCCCG-CDSNMRVQQ-----CETDIREV 917
QY 59 NQ-----SCQCKTIVKRKGSPATPGDKDEGLADEGVFENPQEKEKSERMIGW 110
Db 918 QOLLGTTLNIVCPC-----APKTRSMADSGTENEGSCTGTMWVAVQHPIGT 968
QY 111 HLTTRGKGEENGEPQY-----DKEVSHNHL--PRLTSRQDTSEFSASPERLSVSTI 161

Db 257 -----PPASLKLNL-GTFGFLGLALNSMYRSCKKFLSKWPFRLSSIVASLVLIHV 308
QY 1011 YPELKGLMGRQRNRPPTIVVWSVLASISLSSLWRRIDP 1049
Db 309 FDSLK---PPSQUELVVFVLSFCKSAVPLASVSPV 343

Db 969 QISDDDEEVEDSGYDMVFDIDDSNITHNSLEAQALFNQREADTHY-----ATV 1017
QY 162 AGKRR---LPYSSDYNQ-----SPNRIVDPVGLGNVAK-----ERVG-G 198
Db 1018 QDKRKYLGSPYSPVSPINTIAEAEVESETSPR---LDATIKLTROPKVKRRLFCRELDSG 1074

Db 199 WKMKQ-EKNG-PVSTQAASERG-GVVIDASTDILADEALLNDEAOPLSRKSVISSR 254
Db 1075 YGSEVEAGTGTQVEKHGPENGDDGQBKDTGDRIGEE-----HPEAEAPINS 1123

Db 205 .INPYRMVIMRLVILCFLPHYRITNPVNNAFALWLWSVT-----CEIWFAWSW--- 302

Db 1124 VREHAGTAGILELIKCKDLRAALLGKFKCFGSLFIDLIRPKSDKTCLDWVAGFGIH 1183

Db 303 --ILQDFPKW-FVNRETYLDRLA-----LRYREGREGESOLA--AVDIFSVFD 346

Db 1184 HSIEAFQOLIEPLSYAIIQWLTNAWNKVLVLLRFKVNKSRTVARTLATLNLIPENQ 1243

QY 347 PLKFPPLV-----TANTVLSILAVDYPDKVSCYVFDGAMLSE--- 387

Db 1244 MLEPKIQSGVALWYFRTGISNASTIVG---EAPEWITROTUVIHLGABSOFLITEM 1299

QY 388 -----SIAETSBFARKWVFPCKYSIEPRADPEWWFAAKTIDLYDKVQTSFVKDRRANK 440

Db 1300 VOWAYNDICEESEA---FEYAOGRGDFDSNARA-----FLNSNMQAKVVDCAATMC 1348

QY 441 REEFEFKIRNALVSKALKCPERGWMQDGTPHPGPNNGDHPGMIOFLGQN 492

Db 1349 RHYKHAEM-----KMSIKQWIKRGSKTEG-TGNWKQPIVQFTRHQN 1389

Search completed: June 16, 2003, 11:08:57
Job time : 25 secs

Copyright (c) 1993 - 2003 Compugen Ltd.

On protein - protein search, using sw model

Run on : June 16, 2003, 11:01:28 ; Search time 45 Seconds
(without alignments)

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Title: AAC39336

Perfect score: 5677

Sequence: 1 MESSGETACKPMKNNIVPQTC.....IDPFITSRVNGPDILECGINC 1065

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

RESULT 1

ID AAW33819 standard; Protein: 1065 AA.

XX AAW33819;

XX DT 06-JUL-1998 (first entry)

DE Arabidopsis cellulose biosynthetic protein Ath-B.

XX KW Celulose; cellulose synthase; RSW1 protein; beta-1,4-glucan; transgenic plant.

XX OS Arabidopsis thaliana var. Columbia.

XX PN WO9800549-A1.

XX PD 08-JAN-1998.

XX PF 24-JUN-1997; 97WO-AU00402.

XX PR 27-JUN-1996; 96AU-0000699.

XX PA (CSIR) COMMONWEALTH SCI & IND RES ORG.

PA (AUSU) UNIV AUSTRALIAN NAT.

XX PI Arioli A, Betzner AS, Peng L, Williamson RE;

XX DR WPI; 1998-086974/08.

N-PSDB; AAV06567.

XX PT DNA encoding cellulose biosynthetic enzyme - useful for manipulation of cellulose and beta-1,4-glucan

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	5655	99.6	1065	19 AAW33819
2	5654	99.6	1065	21 AAG4873
3	5654	99.6	1065	23 ABB93356
4	5597	98.6	1054	21 AAG48734
5	5066	89.2	959	21 AAG48735
6	4572	80.5	1077	21 AAY84108
7	4572	80.5	1077	21 AAY84114
8	4564	80.4	1079	23 AAU1096
9	4529.5	79.8	1076	21 AAY84121
10	4418	77.8	1043	21 AAY84119

ALIGNMENTS

11	4028.5	71.0	1075	21 AAY84109	Amino acid sequenc
12	4028.5	71.0	1075	21 AAY84115	Amino acid sequenc
13	4028.5	71.0	1075	21 AAY84120	Amino acid sequenc
14	4024	70.9	1074	21 AAY84110	Amino acid sequenc
15	4024	70.9	1074	21 AAY84112	Amino acid sequenc
16	4024	70.9	1074	21 AAY84117	Amino acid sequenc
17	3984	70.2	1081	19 AAW33817	Arabidopsis celluli
18	3984	70.2	1081	23 ABB93199	Arabidopsis celluli
19	3974	70.0	1081	19 AAW33820	Arabidopsis celluli
20	3922	69.1	1148	21 AAY84118	Corn cellulose syn
21	3900	68.7	1088	21 AAY84107	Corn cellulose syn
22	3875.5	68.3	1086	21 AAY84111	Amino acid sequenc
23	3875.5	68.3	1086	21 AAY84116	Amino acid sequenc
24	3875.5	68.3	1086	21 AAY84135	Corn cellulose syn
25	3843.5	67.7	1094	21 AAY84118	Amino acid sequenc
26	3843.5	67.7	1094	23 ABB93701	Amino acid sequenc
27	3843.5	67.7	1165	21 AYV58834	Corn cellulose syn
28	3811	67.1	1065	23 ABB93522	Herbically activ
29	3804.5	67.0	1039	21 AAY58837	Soybean cellulose
30	3801	67.0	1065	23 ABB91880	Herbically activ
31	3748.5	66.7	1039	19 AAW73309	Corn cellulose syn
32	3741	65.9	1043	23 ABB93701	Cellulose synthase
33	3670	64.6	1069	23 ABB93411	Herbically activ
34	3647.5	64.3	1084	23 ABB93304	Arabidopsis thalia
35	3628	63.9	1084	22 AAB37895	Arabidopsis thalia
36	3596	63.3	1080	22 AAB37895	Arabidopsis thalia
37	3596	63.3	1084	19 AAW33818	Arabidopsis thalia
38	3590	63.2	1088	23 ABB91831	Arabidopsis thalia
39	3522	62.0	821	23 AAU10495	Herbically activ
40	3475.5	61.1	974	19 AAW60039	Cotton cellulose
41	3473.5	61.2	974	19 AAW60039	Eucalyptus grandis
42	3338	59.9	955	21 AAB16338	Populus tremuloide
43	3317	58.4	958	23 AYV58839	Populus tremuloide
44	3216	56.6	793	21 AYV58839	Soybean cellulose
45	3143	55.4	946	22 AAB37893	Populus tremuloide

xx Claim 29; Page 144-149; 207pp; English.
 PS
 XX
 CC This polypeptide is encoded by cDNA clone Ath-B (see AAV06567) that was isolated from an Arabidopsis thaliana cDNA library using PCR primers (see NAT98632-34) based on cellulose synthase RSW1 genomic clone 23H12 (see AAV06563) and EST clone AAT20782 (see AAV06562). It is closely related to Arabidopsis cellulose synthase RSW1 protein (see AAW3816-17). Claimed nucleic acid molecules (see AAV06522-69) coding for claimed polypeptides (see AAW3816-20 and AAW46202) involved in cellulose biosynthesis can be used to manipulate the cellulose and/or beta-glucan content of transgenic plants. Expression of nucleic acids in the sense orientation increases the level of cellulose and reduces the level of non-crystalline beta-1,4-glucan and starch, providing plants with modified strength and/or shape and/or fibre properties, or having increased resistance to stresses or pests. Antisense, ribozyme or co-suppression molecules can be used to reduce the cellulose content of a transgenic plant, e.g. to improve digestibility or to alter carbon partitioning such that increased carbon is available for growth, rather than deposited as cellulose.

CC Sequence 1065 AA;

Query Match 99.6%; Score 5655; DB 19; length 1065;

Best Local Similarity 99.7%; Pred. 0; Mismatches 0; Indels 0; Gaps 0; Matches 1062; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MESEGETAGKPMNIVPOTCQCSDNYGKTVDDREVAACIDCSFPVCPCYEVERKGQNQ 60

Db 1 |||||||MESEGETAGKPMNIVPOTCQCSDNYGKTVDDREVAACIDCSFPVCPCYEVERKGQNQ 60

QY 61 SCPOCKTYKRKGSPATPGDKDKEGLADEGTWFENYFQEKKI-SERMGWHLTRKGEM 120

Db 61 |||||||SCPOCKTYKRKGSPATPGDKDKEGLADEGTWFENYFQEKKI-SERMGWHLTRKGEM 120

QY 121 GEQYDKWSHNILPLRISQDTSGERFAASPERLSVSSPIAGGRLLYSSDVNQSNRR 180

Db 121 |||||||GEQYDKWSHNILPLRISQDTSGERFAASPERLSVSSPIAGGRLLYSSDVNQSNRR 180

QY 181 TIVPVGLGNWANWTERVDCWKMKOKEKNFTGPFVSTOAASERGGVTDASTDILABALLDEA 240

Db 181 |||||||TIVPVGLGNWANWTERVDCWKMKOKEKNFTGPFVSTOAASERGGVTDASTDILABALLDEA 240

QY 241 RQPLSRKVSISSRINPRYMIMLRVILCLEYHYRITNPVPAFLWLVVICIEIWAL 300

Db 241 |||||||RQPLSRKVSISSRINPRYMIMLRVILCLEYHYRITNPVPAFLWLVVICIEIWAL 300

QY 301 SWILDQFPKPWFPNRETYLDRILALRYDREGESQLAANDIFVSTDPIKEPPVANTVIL 360

Db 301 |||||||SWILDQFPKPWFPNRETYLDRILALRYDREGESQLAANDIFVSTDPIKEPPVANTVIL 360

QY 361 SILAVDYPDVKSYCYVFDDGAAMLFSSELAETSEFARKWPKRCKYSIPEPAPEWYRAK 420

Db 361 |||||||SILAVDYPDVKSYCYVFDDGAAMLFSSELAETSEFARKWPKRCKYSIPEPAPEWYRAK 420

QY 421 IDYLKDQVTSFYKVRDRAMKREVEEKFIRKRNALVSKALKCPEEGWMDGTPEWGNIGD 480

Db 421 |||||||IDYLKDQVTSFYKVRDRAMKREVEEKFIRKRNALVSKALKCPEEGWMDGTPEWGNIGD 480

QY 481 HPMQIOFLQGONGLDAEGNELEPRLVVSREKPGFHAKKAGMNLVRSVAVLNGP 540

Db 481 |||||||HPMQIOFLQGONGLDAEGNELEPRLVVSREKPGFHAKKAGMNLVRSVAVLNGP 540

QY 541 ILMUDCDHYINNSKALRAMECFMLDPNIGKQCYVQFOPDFCIDKNDRYANNTVFDI 600

Db 541 |||||||ILMUDCDHYINNSKALRAMECFMLDPNIGKQCYVQFOPDFCIDKNDRYANNTVFDI 600

QY 541 ILMUDCDHYINNSKALRAMECFMLDPNIGKQCYVQFOPDFCIDKNDRYANNTVFDI 600

Db 541 |||||||ILMUDCDHYINNSKALRAMECFMLDPNIGKQCYVQFOPDFCIDKNDRYANNTVFDI 600

QY 601 NURGLDGIGGPVYVGTCGVNRFLALYGYEPPIKVKHKPSLUSKLCCGSRKNSKAKES 660

Db 601 |||||||NURGLDGIGGPVYVGTCGVNRFLALYGYEPPIKVKHKPSLUSKLCCGSRKNSKAKES 660

QY 661 DKKSGRHTDSTVFVNLDIEGVEAGGFDEKALLMSOMSLERFGQSAVFVASTLME 720

Db 661 |||||||DKKSGRHTDSTVFVNLDIEGVEAGGFDEKALLMSOMSLERFGQSAVFVASTLME 720

QY 721 NGCPVPSAPENLNLKEATHVISGVEDSDWGMEIGWYGSVEDILGFKMARGWASI 780

Db 721 |||||||NGCPVPSAPENLNLKEATHVISGVEDSDWGMEIGWYGSVEDILGFKMARGWASI 780

QY 781 YCMPLKPAFKGSAPINLSDLNQVLWALGSVETLFSRCPWYNGRLKEFRAVYN 840

Db 781 |||||||YCMPLKPAFKGSAPINLSDLNQVLWALGSVETLFSRCPWYNGRLKEFRAVYN 840

QY 841 TTIVPITSIPPLMWCFLAVCLFTNQFTIPOISNIASTWFLSFLSIFATGILEMMRSGV 900

Db 841 |||||||TTIVPITSIPPLMWCFLAVCLFTNQFTIPOISNIASTWFLSFLSIFATGILEMMRSGV 900

QY 901 GIDMWWRNEQFWYIGVGAHLAEFVFOGTLKVLAGIDTNTVTKASADSDGDRKELYLTKW 960

Db 901 |||||||GIDMWWRNEQFWYIGVGAHLAEFVFOGTLKVLAGIDTNTVTKASADSDGDRKELYLTKW 960

QY 961 TTLLIPPTLLIVLWLGWAGVSYAINSGYOSQNLPEKLFKAFFWVIVHLYPLKGLMR 1020

Db 961 |||||||TTLLIPPTLLIVLWLGWAGVSYAINSGYOSQNLPEKLFKAFFWVIVHLYPLKGLMR 1020

QY 1021 QNRFTIVVVWSVILLASIFSLWNRIDPFTRVIGPDILECGINC 1065

Db 1021 |||||||QNRFTIVVVWSVILLASIFSLWNRIDPFTRVIGPDILECGINC 1065

RESULT 2
 AR48733 standard; Protein: 1065 AA.
 AR48733:
 XX DT 18-OCT-2000 (first entry)
 XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 61573.
 XX KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX OS Arabidopsis thaliana.
 XX PN EP1033405-A2.
 XX PD 06-SEP-2000.
 XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.
 PR 03-MAR-1999; 99US-0123180.
 PR 03-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 23-MAR-1999; 99US-0126264.
 PR 23-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-012462.
 PR 08-APR-1999; 99US-0128234.
 PR 16-APR-1999; 99US-0128714.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.

PR	14-MAY-1999;	99US-0134218.	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	99US-0134219.	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	99US-0134221.	PR	27-JUL-1999;	99US-0145913.
PR	14-MAY-1999;	99US-0134370.	PR	27-JUL-1999;	99US-0145918.
PR	18-MAY-1999;	99US-0134768.	PR	27-JUL-1999;	99US-0145919.
PR	19-MAY-1999;	99US-0134941.	PR	28-JUL-1999;	99US-0145951.
PR	20-MAY-1999;	99US-0135124.	PR	02-AUG-1999;	99US-0146386.
PR	21-MAY-1999;	99US-0135353.	PR	02-AUG-1999;	99US-0146388.
PR	24-MAY-1999;	99US-0135629.	PR	02-AUG-1999;	99US-0146389.
PR	25-MAY-1999;	99US-0136021.	PR	03-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	99US-0136392.	PR	04-AUG-1999;	99US-0147204.
PR	28-MAY-1999;	99US-0136782.	PR	04-AUG-1999;	99US-0147302.
PR	01-JUN-1999;	99US-0137222.	PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	99US-0137722.	PR	06-AUG-1999;	99US-0147341.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0138847.	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	99US-0139119.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-014929.
PR	18-JUN-1999;	99US-0139454.	PR	15-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139462.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139463.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139464.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139465.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139750.	PR	23-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139763.	PR	23-AUG-1999;	99US-0149902.
PR	21-JUN-1999;	99US-0139817.	PR	17-AUG-1999;	99US-0149175.
PR	22-JUN-1999;	99US-0139899.	PR	18-AUG-1999;	99US-0149426.
PR	23-JUN-1999;	99US-0140353.	PR	20-AUG-1999;	99US-0149722.
PR	24-JUN-1999;	99US-0140354.	PR	20-AUG-1999;	99US-0149723.
PR	24-JUN-1999;	99US-0140695.	PR	20-AUG-1999;	99US-0149929.
PR	28-JUN-1999;	99US-0139750.	PR	23-AUG-1999;	99US-015070.
PR	06-JUL-1999;	99US-0140823.	PR	13-AUG-1999;	99US-0151080.
PR	07-JUL-1999;	99US-0140991.	PR	15-AUG-1999;	99US-0151303.
PR	09-JUN-1999;	99US-0141287.	PR	31-AUG-1999;	99US-0151438.
PR	01-JUL-1999;	99US-0141842.	PR	01-SEP-1999;	99US-0151930.
PR	02-JUL-1999;	99US-0142154.	PR	07-SEP-1999;	99US-0152363.
PR	14-JUL-1999;	99US-0142055.	PR	10-SEP-1999;	99US-015273.
PR	15-JUL-1999;	99US-0142390.	PR	27-AUG-1999;	99US-015306.
PR	16-JUL-1999;	99US-0142803.	PR	27-AUG-1999;	99US-015306.
PR	16-JUL-1999;	99US-0142920.	PR	30-AUG-1999;	99US-0153303.
PR	17-JUL-1999;	99US-0142977.	PR	25-AUG-1999;	99US-015366.
PR	13-JUL-1999;	99US-0143542.	PR	01-SEP-1999;	99US-0154779.
PR	14-JUL-1999;	99US-0143624.	PR	22-SEP-1999;	99US-0155139.
PR	15-JUL-1999;	99US-0143605.	PR	23-SEP-1999;	99US-0155486.
PR	16-JUL-1999;	99US-0144085.	PR	24-SEP-1999;	99US-0155659.
PR	16-JUL-1999;	99US-0144085.	PR	15-SEP-1999;	99US-0156018.
PR	19-JUL-1999;	99US-0144335.	PR	16-SEP-1999;	99US-0156039.
PR	20-JUL-1999;	99US-0144325.	PR	20-SEP-1999;	99US-0156596.
PR	19-JUL-1999;	99US-0144331.	PR	04-OCT-1999;	99US-0157117.
PR	19-JUL-1999;	99US-0144332.	PR	05-OCT-1999;	99US-0157753.
PR	19-JUL-1999;	99US-0144333.	PR	06-OCT-1999;	99US-015865.
PR	19-JUL-1999;	99US-0144334.	PR	07-OCT-1999;	99US-0158029.
PR	19-JUL-1999;	99US-0144335.	PR	08-OCT-1999;	99US-0158458.
PR	20-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	20-JUL-1999;	99US-0144632.	PR	13-OCT-1999;	99US-0159293.
PR	20-JUL-1999;	99US-0144884.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144884.	PR	13-OCT-1999;	99US-0159295.
PR	21-JUL-1999;	99US-0145086.	PR	14-OCT-1999;	99US-0159329.
PR	21-JUL-1999;	99US-0145086.	PR	14-OCT-1999;	99US-0159330.
PR	22-JUL-1999;	99US-014505.	PR	14-OCT-1999;	99US-0159331.
PR	22-JUL-1999;	99US-0145087.	PR	27-JUL-1999;	99US-0145918.
PR	22-JUL-1999;	99US-0145089.	PR	27-JUL-1999;	99US-0145919.
PR	22-JUL-1999;	99US-0145192.	PR	27-JUL-1999;	99US-0145951.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0146386.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0146388.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0146389.

QY	1	MESEGETAKPMKNIVPQTQCICSDNVGKTVGDGRFVACDICSFPCRPCYERKGNO	99.6%; Score 5654; DB 21; Length 1065; Best Local Similarity 99.7%; Pred. No. 0; Mismatches 0; Indels 3; Gaps 0;	Db	1021 QNRPPTIVWWSVILLASIFSLMWRIDPFSRVTGPDILECGINC 1065
Db		1 MESEGETAKPMKNIVPQTQCICSDNVGKTVGDGRFVACDICSFPCRPCYERKGNO	99.6%; Score 5654; DB 21; Length 1065; Best Local Similarity 99.7%; Pred. No. 0; Mismatches 0; Indels 3; Gaps 0;	Db	841 TTYPITSPILIMCYTLPAVCLFTNQFIPQISNIASIWELSFLSIFATGLMRNSGV 900
PR	22-OCT-1999;	99US-0160989.			
PR	25-OCT-1999;	99US-0161404.			
PR	25-OCT-1999;	99US-0161405.			
PR	26-OCT-1999;	99US-0161359.			
PR	26-OCT-1999;	99US-0161360.			
PR	26-OCT-1999;	99US-0161361.			
PR	28-OCT-1999;	99US-0161920.			
PR	28-OCT-1999;	99US-0161992.			
PR	28-OCT-1999;	99US-0161993.			
PR	29-OCT-1999;	99US-0162142.			
Query Match	99.6%; Score 5654; DB 21; Length 1065; Best Local Similarity 99.7%; Pred. No. 0; Mismatches 0; Indels 3; Gaps 0;	Db	901 GIDBWWRNEOFWVIGVGAHLFAVEQGILVLAGIDNTVSKASBDGDAEYLAFK 960		
Matches 1062; Conservative		Db	901 GIDBWWRNEOFWVIGVGAHLFAVEQGILVLAGIDNTVSKASBDGDAEYLAFK 960		
QY	61 SCPOCKTRYKRLKGSPAIFGDKDDEGLADEGTVEFNYPQEKEIISERMIGWHLTRGKGEEM	120			
Db	61 SCFOCKTRYKRLKGSPAIFGDKDDEGLADEGTVEFNYPQEKEIISERMIGWHLTRGKGEEM	120			
QY	121 GEPOYDKEVSHNLPLRSQDTGEFSASPERLVSSTIAGKRLPYSSDVNQSPRR	180			
Db	121 GEPOYDKEVSHNLPLRSQDTGEFSASPERLVSSTIAGKRLPYSSDVNQSPRR	180			
QY	181 IVPVGILGNVAKWERGVSKWMKQEKNTGPESTOASERGVDDASTDILADEALLNDEA	240			
Db	181 IVPVGILGNVAKWERGVSKWMKQEKNTGPESTOASERGVDDASTDILADEALLNDEA	240			
QY	301 SWILDQFPFWPYNPNEYTIDRLALRYDREGPSOLAATIEVSTDPUKEPLVANTWL	360			
Db	301 SWILDQFPFWPYNPNEYTIDRLALRYDREGPSOLAATIEVSTDPUKEPLVANTWL	360			
QY	361 SILADVDPDVKVSQYFDDGAMLSFESELAETSEFARKWVPECKYSTEPRAPEWYRAK	420			
Db	361 SILADVDPDVKVSQYFDDGAMLSFESELAETSEFARKWVPECKYSTEPRAPEWYRAK	420			
QY	421 IDYLDKQVTSFKDRRAMKREVEEFKIRINALVSKALKCPEEGVWMMODGTPAPGNNGD	480			
Db	421 IDYLDKQVTSFKDRRAMKREVEEFKIRINALVSKALKCPEEGVWMMODGTPAPGNNGD	480			
QY	481 HPGMLOQVFGONGLDAKGNELRLVYVSYREKFGFOHKKACAMNALVRSAVLNGPF	540			
Db	481 HPGMLOQVFGONGLDAKGNELRLVYVSYREKFGFOHKKACAMNALVRSAVLNGPF	540			
QY	541 ILNLDCDHYINNSKALREMCFLMDPNIGKQVYQVFOPREGDIDKNDRYYANNTVFDI	600			
Db	541 ILNLDCDHYINNSKALREMCFLMDPNIGKQVYQVFOPREGDIDKNDRYYANNTVFDI	600			
QY	601 NLRGDGQGPVYVGTGCVENRPLALYGEPPPIVKHKPFSSLSKLCGGSSRKNSKAKKES	660			
Db	601 NLRGDGQGPVYVGTGCVENRPLALYGEPPPIVKHKPFSSLSKLCGGSSRKNSKAKKES	660			
QY	661 DKKSGSRHDSTPVFVNLDIERVEGAGFDDERKALLMSOMSLERKFQGSAAVASTME	720			
Db	661 DKKSGSRHDSTPVFVNLDIERVEGAGFDDERKALLMSOMSLERKFQGSAAVASTME	720			
QY	721 NGVPPSAMPAENLKEAIHVISGYEDISDWGMHEIGWIGWYGSWEDILIGFKMARGRSI	780			
Db	721 NGVPPSAMPAENLKEAIHVISGYEDISDWGMHEIGWIGWYGSWEDILIGFKMARGRSI	780			
QY	781 YCMPLKPKFGSAPINSLDRNLQVRLWALGSVTEFLSRHKPIWGYNGRLKELFERAYWN	840			
Db	781 YCMPLKPKFGSAPINSLDRNLQVRLWALGSVTEFLSRHKPIWGYNGRLKELFERAYWN	840			
QY	841 TTYPITSPILIMCYTLPAVCLFTNQFIPQINSIASIWFLSLFLSTFATGILEMRNSGV	900			

Db	61 SCPQCKTRYKRLGSPATPGDKDEGLADEGVVEFPQKEITSERMLGWHLTRGGCEM 120
Oy	121 GEPQYKEVSHNHLPLRTSROTSGEFAASPERLVSSTIAGKRLPYSSDVQSPNRR 180
Db	121 GEPQYKEVSHNHLPLRTSROTSGEFAASPERLVSSTIAGKRLPYSSDVQSPNRR 180
181	IYDVGIGLNWANKERKGQWKMQEKTGPVSTQAASERGGVUDASDILADEALINDEA 240
Db	181 IYDVGIGLNWANKERKGQWKMQEKTGPVSTQAASERGGVUDASDILADEALINDEA 240
Oy	241 ROPLSRKVISSSRINITYRMVIMLRVILCFLHYRITNPVNFAIWLVSVCTWFAL 300
Db	241 ROPLSRKVISSSRINITYRMVIMLRVILCFLHYRITNPVNFAIWLVSVCTWFAL 300
Qy	301 SWLDFQPWKWFPVNRETYLDLRLAKRUREGEPSQLAAVDIFYSTVDPLKEPPVNTVL 360
Db	301 SWLDFQPWKWFPVNRETYLDLRLAKRUREGEPSQLAAVDIFYSTVDPLKEPPVNTVL 360
Oy	361 SILAVDYPDKSYCYFDDGAMLSFSLAETSEFARKWVPPCKKSIETPRAPEWFAAK 420
Db	361 SILAVDYPDKSYCYFDDGAMLSFSLAETSEFARKWVPPCKKSIETPRAPEWFAAK 420
Qy	421 IDYLKDVKVQTSEVKDRRAKREYEEFKIRINALVSKALKCPREGWWMDGTTPWPGNTRD 480
Db	421 IDYLKDVKVQTSEVKDRRAKREYEEFKIRINALVSKALKCPREGWWMDGTTPWPGNTRD 480
Oy	481 HPGMIOVFLGONGLDAEGNELPLRVVSREKRPGFOHKKAGAMNALVRYSAVLNGPF 540
Db	481 HPGMIOVFLGONGLDAEGNELPLRVVSREKRPGFOHKKAGAMNALVRYSAVLNGPF 540
Oy	541 IUNLDCDHYINNSKALRREAMCFLMDPMLPGKQCVYVFQRFIDKNDRYANRNTFFDI 600
Db	541 IUNLDCDHYINNSKALRREAMCFLMDPMLPGKQCVYVFQRFIDKNDRYANRNTFFDI 600
Qy	601 NURGLDGTQGPVYVGTCVFNRLTALGYPEPPRKVKHKRKPSSLKLGGSRRKSKAKES 660
Db	601 NURGLDGTQGPVYVGTCVFNRLTALGYPEPPRKVKHKRKPSSLKLGGSRRKSKAKES 660
Oy	661 DKKSGRHTDSTVPVNLDLLEGGVEGAGFDEDEKALMSQMSLEKRGQSAYVASTLME 720
Db	661 DKKSGRHTDSTVPVNLDLLEGGVEGAGFDEDEKALMSQMSLEKRGQSAYVASTLME 720
Oy	721 NGGVPSATPENLKLKEIHWIISCGYEDKSDWMEIISGMIGSYTEDIITGFKHAROWRSI 780
Db	721 NGGVPSATPENLKLKEIHWIISCGYEDKSDWMEIISGMIGSYTEDIITGFKHAROWRSI 780
Qy	781 YCMKPLKAFKGSAPINLSEDRNLNQVRLWALGSVEILSRHCPCWYGNGRLKLERAYN 840
Db	781 YCMKPLKAFKGSAPINLSEDRNLNQVRLWALGSVEILSRHCPCWYGNGRLKLERAYN 840
Oy	841 TTYPTPSIPLMCLTLLAVCLETNQFIPQISNIASIWFLSIFATGLEMMRSGV 900
Db	841 TTYPTPSIPLMCLTLLAVCLETNQFIPQISNIASIWFLSIFATGLEMMRSGV 900
Oy	901 GIDEWMRNEQFWIGVYSAHLFAVFGQILKVLAGIINTFTVSKASDDEGDFAELYLFWK 960
Db	901 GIDEWMRNEQFWIGVYSAHLFAVFGQILKVLAGIINTFTVSKASDDEGDFAELYLFWK 960
Qy	961 TTYLIPPTTLLVNLWVWAGVYSAHNGYQSWGPFGKLFAPFWVWVHYPFLKGIMGR 1020
Db	961 TTYLIPPTTLLVNLWVWAGVYSAHNGYQSWGPFGKLFAPFWVWVHYPFLKGIMGR 1020
Oy	1021 ONRTPTIVVWSVLLASIFSLWWRDPFTSRVTGPDILEGGINC 1065
Db	1021 ONRTPTIVVWSVLLASIFSLWWRDPFTSRVTGPDILEGGINC 1065

DT	DE	Arabidopsis thaliana protein	18-CCT-2000 (first entry)
XX	XX	Protein identification; signifi-	
KW	KW	cant identification assay; genetic id-	
KW	KW	termination sequence.	
Arabidopsis thaliana.			
XX	OS	Arabidopsis thaliana.	
XX	OS	Arabidopsis thaliana.	
PN	EP1033405-A2.		
XX	PD	06-SEP-2000.	
XX	PD	25-FEB-2000; 2000EP-0301439.J	
PR	PR	25-FEB-1999; 9905-0121825	
PR	PR	05-MAR-1999; 9905-0123180	
PR	PR	09-MAR-1999; 9905-0123548	
PR	PR	23-MAR-1999; 9905-0125764	
PR	PR	25-MAR-1999; 9905-0126264	
PR	PR	29-MAR-1999; 9905-0126783	
PR	PR	01-APR-1999; 9905-0127742	
PR	PR	05-APR-1999; 9905-0128342	
PR	PR	08-APR-1999; 9905-0128714	
PR	PR	16-APR-1999; 9905-0129845	
PR	PR	19-APR-1999; 9905-0130077	
PR	PR	21-APR-1999; 9905-0130449	
PR	PR	23-APR-1999; 9905-0130509	
PR	PR	23-APR-1999; 9905-0130891	
PR	PR	28-APR-1999; 9905-0131490	
PR	PR	30-APR-1999; 9905-0132048	
PR	PR	30-APR-1999; 9905-0132470	
PR	PR	04-MAY-1999; 9905-0132484	
PR	PR	05-MAY-1999; 9905-0132485	
PR	PR	06-MAY-1999; 9905-0132487	
PR	PR	07-MAY-1999; 9905-0132863	
PR	PR	11-MAY-1999; 9905-0133428	
PR	PR	14-MAY-1999; 9905-0134248	
PR	PR	14-MAY-1999; 9905-0134249	
PR	PR	14-MAY-1999; 9905-0134250	
PR	PR	14-MAY-1999; 9905-0134251	
PR	PR	18-MAY-1999; 9905-0134792	
PR	PR	20-MAY-1999; 9905-0134941	
PR	PR	21-MAY-1999; 9905-0135124	
PR	PR	24-MAY-1999; 9905-0135353	
PR	PR	25-MAY-1999; 9905-0136021	
PR	PR	27-MAY-1999; 9905-0136392	
PR	PR	28-MAY-1999; 9905-0136783	
PR	PR	01-JUN-1999; 9905-0137222	
PR	PR	03-JUN-1999; 9905-0137528	
PR	PR	04-JUN-1999; 9905-0137502	
PR	PR	07-JUN-1999; 9905-0137724	
PR	PR	08-JUN-1999; 9905-0138094	
PR	PR	10-JUN-1999; 9905-0138451	
PR	PR	10-JUN-1999; 9905-0138847	
PR	PR	14-JUN-1999; 9905-0139119	
PR	PR	16-JUN-1999; 9905-0139522	
PR	PR	16-JUN-1999; 9905-0139453	
PR	PR	17-JUN-1999; 9905-0139459	
PR	PR	18-JUN-1999; 9905-0139541	
PR	PR	18-JUN-1999; 9905-0139455	
PR	PR	18-JUN-1999; 9905-0139461	
PR	PR	18-JUN-1999; 9905-0139462	
PR	PR	18-JUN-1999; 9905-0139463	
PR	PR	18-JUN-1999; 9905-0139750	

Query	Match	Score	DB	Length
	Best Local Similarity	98.6%	21	1054
	Matches	99.7%	0	
	Matches	0	0	
	Conservative	0	0	
	Mismatches	3	0	
	Indels	0	0	
	Gaps	0	0	
Qy	12 MKNIVPQTCQICSDNVGKUWGDPRFVACDICSCPVCPYCYEYERKDENOSCPOCKTRYKR	71		
Db	1 MKNIVPQTCQICSDNVGKUWGDPRFVACDICSCPVCPYCYEYERKDENOSCPOCKTRYKR	60		
Qy	72 LKSPPAIOPDKDGLADEGTVFENPYKEKISERMUGWHLTRGKGEMGEPOYDKVSH	131		
Db	61 LKSPPAIOPDKDGLADEGTVFENPYKEKISERMUGWHLTRGKGEMGEPOYDKVSH	120		
Qy	132 NHILPRLTSRQDTSGEFSASPERLUSVSTIAGKRLPPSYSDYNQSPNRIRVPVGJNVA	191		
Db	121 NHILPRLTSRQDTSGEFSASPERLUSVSTIAGKRLPPSYSDYNQSPNRIRVPVGJNVA	180		
Qy	192 WKERVDSKMKQEKNTGKVSTOASERGKVDTADSTPLADEFALNDEAROPLSKVSYIP	251		
Db	181 WKERVDSKMKQEKNTGKVSTOASERGKVDTADSTPLADEFALNDEAROPLSKVSYIP	240		
Qy	252 SSINPYKWMIRLVLICFLHYRTINPVPNAFLMUVSICEIWALMSWLDQFQKWF	311		
PR	18-JUN-1999; 99US-0139763.		PR	27-AUG-1999; 99US-0151066.
PR	21-JUN-1999; 99US-0139817.		PR	27-AUG-1999; 99US-0151080.
PR	22-JUN-1999; 99US-0139893.		PR	30-AUG-1999; 99US-0151303.
PR	23-JUN-1999; 99US-0140353.		PR	31-AUG-1999; 99US-0151438.
PR	24-JUN-1999; 99US-0140695.		PR	01-SEP-1999; 99US-0151930.
PR	28-JUN-1999; 99US-0140823.		PR	07-SEP-1999; 99US-0152363.
PR	29-JUN-1999; 99US-0140991.		PR	10-SEP-1999; 99US-0153070.
PR	30-JUN-1999; 99US-0141287.		PR	13-SEP-1999; 99US-0153758.
PR	01-JUL-1999; 99US-0141842.		PR	15-SEP-1999; 99US-0154018.
PR	01-JUL-1999; 99US-0142154.		PR	16-SEP-1999; 99US-0154039.
PR	02-JUL-1999; 99US-0142055.		PR	20-SEP-1999; 99US-0154779.
PR	06-JUL-1999; 99US-0142390.		PR	22-SEP-1999; 99US-0155139.
PR	08-JUL-1999; 99US-0142803.		PR	23-SEP-1999; 99US-0155486.
PR	09-JUL-1999; 99US-0142920.		PR	24-SEP-1999; 99US-0155659.
PR	12-JUL-1999; 99US-0142977.		PR	28-SEP-1999; 99US-0156458.
PR	13-JUL-1999; 99US-0143542.		PR	29-SEP-1999; 99US-0156596.
PR	14-JUL-1999; 99US-0143624.		PR	04-OCT-1999; 99US-0157117.
PR	15-JUL-1999; 99US-0144005.		PR	05-OCT-1999; 99US-0157753.
PR	16-JUL-1999; 99US-0144086.		PR	06-OCT-1999; 99US-0157865.
PR	19-JUL-1999; 99US-0144325.		PR	07-OCT-1999; 99US-0158029.
PR	19-JUL-1999; 99US-0144331.		PR	08-OCT-1999; 99US-0158232.
PR	19-JUL-1999; 99US-0144337.		PR	12-OCT-1999; 99US-0158369.
PR	19-JUL-1999; 99US-0144337.		PR	13-OCT-1999; 99US-0159293.
PR	19-JUL-1999; 99US-0144337.		PR	14-OCT-1999; 99US-0159294.
PR	19-JUL-1999; 99US-0144337.		PR	13-OCT-1999; 99US-0159295.
PR	21-JUL-1999; 99US-0145086.		PR	14-OCT-1999; 99US-0159329.
PR	19-JUL-1999; 99US-0144335.		PR	21-OCT-1999; 99US-0159330.
PR	20-JUL-1999; 99US-0144335.		PR	14-OCT-1999; 99US-0159331.
PR	20-JUL-1999; 99US-0144632.		PR	14-OCT-1999; 99US-0159637.
PR	20-JUL-1999; 99US-0144884.		PR	14-OCT-1999; 99US-0159638.
PR	21-JUL-1999; 99US-0144884.		PR	18-OCT-1999; 99US-0159584.
PR	21-JUL-1999; 99US-0145192.		PR	21-OCT-1999; 99US-0160741.
PR	21-JUL-1999; 99US-0145195.		PR	22-OCT-1999; 99US-0160980.
PR	21-JUL-1999; 99US-0145218.		PR	22-OCT-1999; 99US-0160981.
PR	23-JUL-1999; 99US-0145224.		PR	22-OCT-1999; 99US-0160989.
PR	22-JUL-1999; 99US-0145287.		PR	25-OCT-1999; 99US-0161040.
PR	22-JUL-1999; 99US-0145897.		PR	25-OCT-1999; 99US-0161045.
PR	22-JUL-1999; 99US-0145913.		PR	25-OCT-1999; 99US-0161406.
PR	27-JUL-1999; 99US-0145918.		PR	26-OCT-1999; 99US-0161359.
PR	27-JUL-1999; 99US-0145919.		PR	26-OCT-1999; 99US-0161360.
PR	28-JUL-1999; 99US-0145951.		PR	26-OCT-1999; 99US-0161361.
PR	26-JUL-1999; 99US-0146386.		PR	28-OCT-1999; 99US-0161920.
PR	27-JUL-1999; 99US-0146389.		PR	28-OCT-1999; 99US-0161992.
PR	03-AUG-1999; 99US-0147038.		PR	28-OCT-1999; 99US-0161993.
PR	04-AUG-1999; 99US-0147204.		PR	29-OCT-1999; 99US-0162142.
PR	04-AUG-1999; 99US-0147303.			
PR	05-AUG-1999; 99US-0147260.			
PR	06-AUG-1999; 99US-0147303.			
PR	06-AUG-1999; 99US-0147416.			
PR	09-AUG-1999; 99US-0147433.			
PR	10-AUG-1999; 99US-0147935.			
PR	11-AUG-1999; 99US-0148319.			
PR	12-AUG-1999; 99US-0148341.			
PR	13-AUG-1999; 99US-0148365.			
PR	13-AUG-1999; 99US-0148684.			
PR	16-AUG-1999; 99US-0149902.			
PR	17-AUG-1999; 99US-0149368.			
PR	18-AUG-1999; 99US-0149416.			
PR	20-AUG-1999; 99US-0149722.			
PR	20-AUG-1999; 99US-0149723.			
PR	23-AUG-1999; 99US-0149902.			
PR	23-AUG-1999; 99US-0149930.			
PR	25-AUG-1999; 99US-0150566.			
PR	26-AUG-1999; 99US-0150884.			
PR	27-AUG-1999; 99US-0151085.			

Db	241 SSRINPYRMVIMRLVILCFLHYRITNPVNAFAILWLVSVICELFALSWILDQFPKW	PD	06-SEP-2000.
Qy	312 PVNRETYLDRIALRYREGESQLAAVDIFVSTDPLKEPPLYTAVLTLAVOYPDK	XX	25-FEB-2000; 2000EP-0301439.
Db	301 PVNRETYLDRIALRYREGESQLAAVDIFVSTDPLKEPPLYTAVLTLAVOYPDK	PR	25-FEB-1999; 99US-0121825.
Qy	372 VSCIVYDDGAAMLFESLAETSEFARKWNPCKKSIETPRAPEWPAAKIDYLKDQVTS	PR	05-MAR-1999; 99US-0123180.
Db	361 VSCIVYDDGAAMLFESLAETSEFARKWNPCKKSIETPRAPEWPAAKIDYLKDQVTS	PR	09-MAR-1999; 99US-0123548.
Qy	432 FVKDRRAMKREVEEFKIRTRNALVSKALKCPERGWMDGTWPGNNNGDHGMIOVFLGQ	PR	23-MAR-1999; 99US-0123788.
Db	421 FVKDRRAMKREVEEFKIRTRNALVSKALKCPERGWMDGTWPGNNNGDHGMIOVFLGQ	PR	25-MAR-1999; 99US-0123264.
Qy	492 NGGLDEEGNELPLRVYVSREKRGCFOHKKAGAMNLVRYSAVLINGPFTLNLDQHYN	PR	29-MAR-1999; 99US-0126785.
Db	481 NGGLDEEGNELPLRVYVSREKRGCFOHKKAGAMNLVRYSAVLINGPFTLNLDQHYN	PR	01-APR-1999; 99US-012462.
Qy	552 NSKLREAMCFLMDPNLGKQCYVQFPQRFCIDKNDRYANRNTFFDINURGLDGQGP	PR	05-APR-1999; 99US-0124234.
Db	541 NSKLREAMCFLMDPNLGKQCYVQFPQRFCIDKNDRYANRNTFFDINURGLDGQGP	PR	08-APR-1999; 99US-0128114.
Qy	612 VVGGTCGVNRITALYGEPEPTIKVHKRPLSLISKLCGGSRKNSKAKKEDKKSGRHTS	PR	16-APR-1999; 99US-012945.
Db	601 VVGGTCGVFNKITALYGEPEPTIKVHKRPLSLISKLCGGSRKNSKAKKEDKKSGRHTS	PR	19-APR-1999; 99US-0130077.
Qy	672 TVPVFNLDIEBEGVEGAGFDEKKALMSOMSLERFGQSAVFASLTENGVPPSATPE	PR	21-APR-1999; 99US-0130449.
Db	661 TVPVFNLDIEBEGVEGAGFDEKKALMSOMSLERFGQSAVFASLTENGVPPSATPE	PR	23-APR-1999; 99US-0130510.
Qy	732 NULKEATHVSCGYEDKSDMGMEIGHTYGSYTEDITGKMHARGRSITCKMPKUPAFKG	PR	23-APR-1999; 99US-0130891.
Db	721 NULKEATHVSCGYEDKSDMGMEIGHTYGSYTEDITGKMHARGRSITCKMPKUPAFKG	PR	28-APR-1999; 99US-0132048.
Qy	792 SAPINUSDRNLNQVLRVALGSWEILFSRRCPTWGYGNRKFELERFAVNTIYTISPL	PR	04-MAY-1999; 99US-013224.
Db	781 SAPINUSDRNLNQVLRVALGSWEILFSRRCPTWGYGNRKFELERFAVNTIYTISPL	PR	05-MAY-1999; 99US-0132484.
Qy	852 LMYCTLAVALCIFTNQITIPOINSIASIWLSLFLSIFATGILEMRWNSGGIDEMWNQF	PR	05-MAY-1999; 99US-0132485.
Db	841 LMYCTLAVALCIFTNQITIPOINSIASIWLSLFLSIFATGILEMRWNSGGIDEMWNQF	PR	06-MAY-1999; 99US-0132486.
Qy	912 WVGIGYSAHLFAVFOCILKVLAGIDONFTVSKASDEDGDAELYLFWKWTLLIPPTL	PR	07-MAY-1999; 99US-0132487.
Db	901 WVGIGYSAHLFAVFOCILKVLAGIDONFTVSKASDEDGDAELYLFWKWTLLIPPTL	PR	10-MAY-1999; 99US-0132488.
Qy	972 IVNLVGVWAGSYAINGSYQSGWGLPLGKLFIAFWVVLHLYFLKGJMGQRNRTPTVVW	PR	11-MAY-1999; 99US-0132489.
Db	961 IVNLVGVWAGSYAINGSYQSGWGLPLGKLFIAFWVVLHLYFLKGJMGQRNRTPTVVW	PR	12-MAY-1999; 99US-0132490.
Qy	1032 SYLLASIFSLLWVRIDPFTSRVTGPDPILECGINC 1065	PR	13-MAY-1999; 99US-0132491.
Db	1021 SYLLASIFSLLWVRIDPFTSRVTGPDPILECGINC 1054	PR	14-MAY-1999; 99US-0132492.
RESULT 5		PR	15-MAY-1999; 99US-0132493.
AAG48735		PR	16-MAY-1999; 99US-0132494.
ID AAG48735 standard; Protein; 959 AA.		PR	17-MAY-1999; 99US-0132495.
XX		PR	18-MAY-1999; 99US-0132496.
AC AAG48735;		PR	19-MAY-1999; 99US-0132497.
XX		PR	20-MAY-1999; 99US-0132498.
DT 18-OCT-2000 (first entry)		PR	21-MAY-1999; 99US-0132499.
DE Arabidopsis thaliana protein fragment SEQ ID NO: 61575.		PR	22-MAY-1999; 99US-0132500.
KW Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.		PR	23-JUN-1999; 99US-0140353.
OS Arabidopsis thaliana.		PR	24-JUN-1999; 99US-0140354.
XX EP1033405-A2.		PR	28-JUN-1999; 99US-0140595.
XX		PR	29-JUN-1999; 99US-0140991.
OS		PR	30-JUN-1999; 99US-014287.
XX PR 01-JUL-1999; 99US-0140842.		PR	02-JUL-1999; 99US-0142154.
XX		PR	02-JUL-1999; 99US-0142055.

	Query	Match	Score	DB	Length	
PR	06-JUL-1999;	990S-0142390.	5066;	DB	21;	
PR	08-JUL-1999;	990S-0142803.	990S-0155485.	PR	23-SEP-1999;	
PR	09-JUL-1999;	990S-0142920.	990S-0155659.	PR	24-SEP-1999;	
PR	12-JUL-1999;	990S-0142977.	990S-0156458.	PR	28-SEP-1999;	
PR	13-JUL-1999;	990S-0143542.	990S-0156596.	PR	29-SEP-1999;	
PR	14-JUL-1999;	990S-0143624.	990S-0157117.	PR	04-OCT-1999;	
PR	15-JUL-1999;	990S-0144005.	990S-0157753.	PR	05-OCT-1999;	
PR	16-JUL-1999;	990S-0144085.	990S-0158029.	PR	07-OCT-1999;	
PR	16-JUL-1999;	990S-0144086.	990S-0158232.	PR	08-OCT-1999;	
PR	19-JUL-1999;	990S-0144325.	990S-0158369.	PR	12-OCT-1999;	
PR	19-JUL-1999;	990S-0144331.	990S-0159293.	PR	13-OCT-1999;	
PR	19-JUL-1999;	990S-0144332.	990S-0159294.	PR	13-OCT-1999;	
PR	19-JUL-1999;	990S-0144333.	990S-0159295.	PR	13-OCT-1999;	
PR	19-JUL-1999;	990S-0144334.	990S-0159329.	PR	14-OCT-1999;	
PR	19-JUL-1999;	990S-0144335.	990S-0159330.	PR	14-OCT-1999;	
PR	20-JUL-1999;	990S-0144352.	990S-0159331.	PR	14-OCT-1999;	
PR	20-JUL-1999;	990S-0144632.	990S-0159637.	PR	14-OCT-1999;	
PR	20-JUL-1999;	990S-0144884.	990S-0159638.	PR	14-OCT-1999;	
PR	21-JUL-1999;	990S-0144814.	990S-0159584.	PR	18-OCT-1999;	
PR	21-JUL-1999;	990S-0145086.	990S-0160741.	PR	21-OCT-1999;	
PR	21-JUL-1999;	990S-0145218.	990S-0160981.	PR	21-OCT-1999;	
PR	22-JUL-1999;	990S-0145224.	990S-0160989.	PR	22-OCT-1999;	
PR	22-JUL-1999;	990S-0145087.	990S-0160768.	PR	22-OCT-1999;	
PR	22-JUL-1999;	990S-0145089.	990S-0160770.	PR	21-OCT-1999;	
PR	22-JUL-1999;	990S-0145192.	990S-0160814.	PR	21-OCT-1999;	
PR	22-JUL-1999;	990S-0145145.	990S-0160815.	PR	21-OCT-1999;	
PR	23-JUL-1999;	990S-0145218.	990S-0160980.	PR	22-OCT-1999;	
PR	23-JUL-1999;	990S-0145224.	990S-0160981.	PR	22-OCT-1999;	
PR	26-JUL-1999;	990S-0145276.	990S-0160989.	PR	22-OCT-1999;	
PR	27-JUL-1999;	990S-0145913.	990S-0161404.	PR	25-OCT-1999;	
PR	27-JUL-1999;	990S-0145918.	990S-0161405.	PR	25-OCT-1999;	
PR	27-JUL-1999;	990S-0145919.	990S-0161406.	PR	25-OCT-1999;	
PR	28-JUL-1999;	990S-0145951.	990S-0161359.	PR	26-OCT-1999;	
PR	02-AUG-1999;	990S-0146386.	990S-0161360.	PR	26-OCT-1999;	
PR	02-AUG-1999;	990S-0146388.	990S-0161361.	PR	26-OCT-1999;	
PR	02-AUG-1999;	990S-0146389.	990S-0161920.	PR	28-OCT-1999;	
PR	03-AUG-1999;	990S-0147038.	990S-0161992.	PR	28-OCT-1999;	
PR	04-AUG-1999;	990S-0147204.	990S-0161993.	PR	29-OCT-1999;	
PR	05-AUG-1999;	990S-0147192.	990S-0162142.			
	Query	Match	Score	DB	Length	
	Best	Local Similarity	89.2%	DB	21;	
	Matches	956;	Pred. No. 0;	Mismatches	99.7%;	
	Matches	956;	Conservative	0;	Indels 0;	
					Gaps 0;	
QY	09-AUG-1999;	990S-0147493.	107	MIGWHLTRGKGEMGEPMGYDKEYSHNHLPLRTSRODTSGESAAASPERLSVSTIAGKR	166	
PR	10-AUG-1999;	990S-0147495.	1	MIGWHLTRGKGEMGEPMGYDKEYSHNHLPLRTSRODTSGESAAASPERLSVSTIAGKR	60	
PR	11-AUG-1999;	990S-0148171.	QY	167	LPSSDWQSPNRRIVDPGLGNWAKWERGVKGKMKQEKNGPVSQAAASERGGVVIDAS	226
PR	12-AUG-1999;	990S-0148341.	61	LPYSSDVNQSPNRRIVDPGLGNWAKWERGVKGKMKQEKNGPVSQAAASERGGVVIDAS	120	
PR	13-AUG-1999;	990S-0148365.	Db	TDLADEALLNDRARQPLSKRSVSISSSINPFRMIMRLVILCFLHYRTPNPNAFA	286	
PR	16-AUG-1999;	990S-0148684.	QY	227	TDLADEALLNDRARQPLSKRSVSISSSINPFRMIMRLVILCFLHYRTPNPNAFA	286
PR	17-AUG-1999;	990S-0149368.	Db	121 TDLADEALLNDRARQPLSKRSVSISSSINPFRMIMRLVILCFLHYRTPNPNAFA	180	
PR	18-AUG-1999;	990S-0149175.	QY	287 LWLVSVCEIWALSWILDQEPKWFPVNRETYLDRLALRYDRBEGPSLAQADIFTSTVD	346	
PR	20-AUG-1999;	990S-0149426.	Db	181 LWLVSVCEIWALSWILDQEPKWFPVNRETYLDRLALRYDRBEGPSLAQADIFTSTVD	240	
PR	20-AUG-1999;	990S-0149722.	QY	347 PLKEPPMTANTLISLAWDYPDKVSYVFDDGAMLSFESLAESFESAREKWPCKY	406	
PR	20-AUG-1999;	990S-0149723.	Db	121 PLKEPPMTANTLISLAWDYPDKVSYVFDDGAMLSFESLAESFESAREKWPCKY	300	
PR	21-AUG-1999;	990S-0149902.	QY	407 SIEPRPENWFAAKDYLKDVKTSFKDRRAMEFLMDPNQKGQVYQFQRFCIDK	526	
PR	23-AUG-1999;	990S-0149930.	Db	301 SIEPRPENWFAAKDYLKDVKTSFKDRRAMEFLMDPNQKGQVYQFQRFCIDK	526	
PR	25-AUG-1999;	990S-0150566.	QY	467 MDPGPWGNNTGDPGMLOVFLGONGSLDAEGRNEPLRVVSYREKRGFOHHKKAGMN	526	
PR	26-AUG-1999;	990S-0150884.	Db	361 MDPGPWGNNTRHPGMLOVFLGONGSLDAEGRNEPLRVVSYREKRGFOHHKKAGMN	420	
PR	01-SEP-1999;	990S-0151930.	QY	527 ALRVSAVINGPLINLDCDHINNSALREALMEFLMDPNQKGQVYQFQRFCIDK	586	
PR	07-SEP-1999;	990S-0152363.				
PR	27-AUG-1999;	990S-0151066.				
PR	13-SEP-1999;	990S-0153758.				
PR	15-SEP-1999;	990S-0154018.				
PR	16-SEP-1999;	990S-0154039.				
PR	20-SEP-1999;	990S-0154779.				
PR	22-SEP-1999;	990S-0155139.				

Db	421 ALVRVSAAVLTNPFLNLDCHYINNSKALRBAACMCLMPDPLUGKQCVYQVQRFEGIDK	PS	Claim 15; Page 95-97; 119pp; English.
Qy	587 NDRYANRNTVFFDINLRGLDGIGQPYVVGTCVFNRTALGYEPPIRKVKHRSPLSLKC	Xx	The present sequence represents a maize cellulose synthase polypeptide.
Db	481 NDRYANRNTVFFDINLRGLDGIGQPYVVGTCVFNRTALGYEPPIRKVKHRSPLSLKC	CC	The cellulose synthase can be used for the improvement of stalk quality for improved stand or silage. It also provides an increased concentration of cellulose in the pericarp, hardening the kernel and improving its handling ability. The sequences are used to produce transgenic plants and seeds expressing the cellulose synthase. The polynucleotide is used for modulating, preferably increasing, the level of the synthase in a plant cell. The plants are preferably monocots. The polynucleotide is also used as a probe or primer in the detection quantitation or isolation of gene transcripts. The probes are useful in detecting deficiencies in the level of mRNA in screenings for desired transgenic plant, for detecting mutations in the gene, for monitoring upregulation of expression or changes in enzyme activity in screening assays of compounds, for detection of any number of allelic variants of the gene, or for use as molecular markers in plant breeding programs. The isolated nucleic acids of the present invention can also be used for recombinant expression of their encoded polypeptides or for use as immunogens in the preparation and/or screening of antibodies. The proteins can be employed in assays for enzyme agonists or antagonists of enzyme function or for use of immunogens or antigens to obtain antibodies specifically immunoreactive with a protein.
Qy	647 GGSRKNSKAKKESDKKKSGRHTSDVYPPVNLADDIEBEGVEGAGFDEKKALMSQSLKR	CC	
Db	541 GGSRKNSKAKKESDKKKSGRHTSDVYPPVNLADDIEBEGVEGAGFDEKKALMSQSLKR	CC	
Qy	707 FGQSAAVFKVASTLMEENGVPPSATPENLKEATHVLSGGYEDKSDWQMEIGWYGSVTDI	CC	
Db	601 FGQSAAVFKVASTLMEENGVPPSATPENLKEATHVLSGGYEDKSDWQMEIGWYGSVTDI	CC	
Qy	767 LTGFKMHARGWSIYCHMCKPLPAFKGSAPPINSLDRNQVRLWALGSVLEILFRHCPIWGY	CC	
Db	661 LTGFKMHARGWSIYCHMCKPLPAFKGSAPPINSLDRNQVRLWALGSVLEILFRHCPIWGY	CC	
Qy	827 NGRLFELERFAVNTTYPITPSIPLMCTLPAVCLNTQFIPOTSNASWFLS	CC	
Db	721 NGRLFELERFAVNTTYPITPSIPLMCTLPAVCLNTQFIPOTSNASWFLS	CC	
Qy	887 IFATGILEMRWSGVGIDEWWRNEQFWIGGSAHLFAWFOGLIKVLAGIDNFNTVSKAS	CC	
Db	781 IFATGILEMRWSGVGIDEWWRNEQFWIGGSAHLFAWFOGLIKVLAGIDNFNTVSKAS	CC	
Qy	947 DEDGDFAEVLYLKWTMILLIPPTTLNVNLVGVAGYSAINGSYOSWGPLGKLFIAFWV	CC	
Db	841 DEGDFAEVLKFWMILLIPPTTLNVNLVGVAGYSAINGSYOSWGPLGKLFIAFWV	CC	
Qy	1007 IVHLYPKLGMGRQNPPTIVVWSVLLASIFSLWVRIPIFTSRVTPGPILECINC	CC	
Db	901 IVHLYPKLGMGRQNPPTIVVWSVLLASIFSLWVRIPIFTSRVTPGPILECINC	CC	
RESULT 6			
ID	AYA84108	Sequence	1077 AA;
XX	AYA84108 standard; Protein; 1077 AA.		
AC	AYA84108;		
XX			
DT	03-JUL-2000 (first entry)		
DE	Amino acid sequence of a maize cellulose synthase.		
XX			
KW	Maize; cellulose synthase; stalk quality; stand; silage; cellulose; transgenic plant; plant breeding marker.		
XX			
OS	zea mays.		
XX			
FH	Key Location/Qualifiers		
FT	Misc-difference 494 /note= "encoded by MG"		
XX			
PN	WO200009706-A2.		
PD	24-FEB-2000.		
XX			
PF	16-AUG-1999; 99W0-US18760.		
XX			
PR	17-AUG-1998; 98US-0096822.		
XX			
(PION-) PIONEER HI-BRED INT INC.			
XX			
PI	Dhugga KS, Heientjariis TG, Bowen BA, Wang X;		
XX			
DR	WPI; 2000-224343/19.		
XX	N-PSDB, AANZ9949.		
PT	New genes which encode maize cellulose synthase polypeptides in plants and to produce transgenic plants expressing the novel protein		
PT			
Qy	522 AGAMNALVRSVAVLTNPFLNLDCHYINNSKALRBAACMCLMPDPLUGKQCVYQVOPRF	Qy	
Db	462 ERGWVMDGTPWPGNNFGDHGPMQVFLGQNGGLDAEGNELPRLVVSREKPGQHHK	Qy	
Qy	537 AGAMNALVRSVAVLTNPFLNLDCHYINNSKALRBAACMCLMPDPLUGKQCVYQVOPRF	Qy	
Db	582 DGIDRNRYANRNTVFFDINLRGLDGIGQPYVVGCCVFNRITALGYEPPIRKVKKKPSL	Qy	
Db	597 DGIDRNRYANRNTVFFDINLRGLDGIGQPYVVGCCVFNRITALGYEPPI-KOKKKGF	Qy	

QY	LSKLGCSRKRKSKAKKESDKKKRGRHLDSTVNFVNLIDIECSEGAEFDDEKALMSQM	701	CC	The present sequence represents a maize cellulose synthase polypeptide.
Db	LSSICGG-RKKASKSKKGSDKKSKOIKHDSSVVFENLDEEGVEAGFDDERSLMSQM	713	CC	The cellulose synthase can be used for the improvement of stalk quality for improved stand or silage. It also provides an increased concentration of cellulose in the pericarp, hardening the kernel and improving its handling ability. The sequences are used to produce transgenic plants and seeds expressing the cellulose synthase. The polynucleotide is used for modulating, preferably increasing, the level of the synthase in a plant cell. The plants are preferably monocots. The polynucleotide is also used as a probe or primer in the detection, quantitation or isolation of gene transcripts. The probes are useful in detecting deficiencies in the level of mRNA in screenings for desired transgenic plant, for detecting mutations in the gene, for monitoring upregulation of expression or changes in enzyme activity in screening assays of compounds, for detection of any number of allelic variants of the gene, or for use as molecular markers in plant breeding programs. The isolated nucleic acids of the present invention can also be used for recombinant expression of their encoded polypeptides or for use as immunogens in the preparation and/or screening of antibodies. The proteins can be employed in assays for enzyme agonists or antagonists of enzyme function or for use of immunogens or antigens to obtain antibodies specifically immunoreactive with a protein.
QY	702 SLEKRGFSAVFASLTMEGGVPPSATPENLKEIAHWISCVYEDKDWGMETGWTGS	761	CC	
Db	714 SLEKRGFSAAFTASTIMEGGVPPSATPESLKEIAHWISCVYEDKDWGMETGWTGS	773	CC	
QY	762 VIEDILGFKMHARGWISIYCHPKPLAKGSAPINSLDRNQYLRWAIGSVELSRICP	821	CC	
Db	774 VIEDILGFKMHARGWISIYCHPKPLAKGSAPINSLDRNQYLRWAIGSVELSRICP	833	CC	
QY	822 IWFNGSALKFLERAVINTTYKPIITSPLLMCTLNQFQDQISIASTWLF	881	CC	
Db	834 LWFYGGRLKELERAYINTTYPLTSPLLYCILPAICLUTGKFELPEISNFASTWLF	893	CC	
QY	882 SIFLSTIFAMGLLEMRSWCGVIDERWWRNQFWWIGGSVAHLFAVFOGLKVLVLAGIDTNTPV	941	CC	
Db	894 SIFISIRFTGLEMMRWSVGIDEMWRNEQFWIGISAHLFAYFOGLKVLVLAGIDTNTPV	953	CC	
QY	942 TKAASDEGDGFASFLYLFLWTTLIPPTPLLIVWVAGVSAINSYQSNGPLFAGLF	1001	CC	
Db	954 TSKASDEGDGFASFLYLFLWTTLIPPTPLLIVWVAGVSAINSYQSNGPLFAGLF	1013	CC	
QY	1002 FAFWVWILYPLFKLGMQRNQTPTIVWVWSVLAISLSSLWRIDPDTSRVGPDILEC	1061	CC	
Db	1014 FAFWVWILYPLFKLGMQRNQTPTIVWVWSVLAISLSSLWRIDPDTSRVGPDILEC	1073	CC	
RESULT 7				
ID	AAY84114 standard; Protein; 1077 AA.			
XX				
AC	AAY84114;			
XX				
DR	03-JUL-2000 (first entry)			
XX				
DE	Amino acid sequence of a maize cellulose synthase.			
XX				
KW	maize; cellulose synthase; stalk quality; stand; silage; cellulose; transgenic plant; plant breeding marker.			
XX				
OS	zea mays.			
XX				
FH	Key Location/Qualifiers			
FT	Misc-difference 494			
FT	/note= "encoded by MGG"			
XX				
PN	W0200009706-A2.			
XX				
PD	24-FEB-2000.			
XX				
PR	16-AUG-1999; 99WO-US18760.			
XX				
PR	17-AUG-1998; 98US-0096822.			
XX				
PA	(PION-) PIONEER HI-BRED INT INC.			
XX				
PI	Dhuga K.S., Helentjaris T.G., Bowen B.A., Wang X;			
XX				
WPI:	2000-224343719.			
DR	N-PSDB: AA2299512.			
XX				
PT	New genes which encode maize cellulose synthase polypeptides in plants useful for modulating the expression of cellulose synthase in plants and to produce transgenic plants expressing the novel protein			
PT	claim 15; Page 142-144; 119pp; English.			
QY	537 AGAMMALVRYSAVLTGPFITNLDDHYTNISKALREAGNEPLRVYVSRKPSOHKK	521	CC	The cellulose synthase can be used for the improvement of stalk quality for improved stand or silage. It also provides an increased concentration of cellulose in the pericarp, hardening the kernel and improving its handling ability. The sequences are used to produce transgenic plants and seeds expressing the cellulose synthase. The polynucleotide is used for modulating, preferably increasing, the level of the synthase in a plant cell. The plants are preferably monocots. The polynucleotide is also used as a probe or primer in the detection, quantitation or isolation of gene transcripts. The probes are useful in detecting deficiencies in the level of mRNA in screenings for desired transgenic plant, for detecting mutations in the gene, for monitoring upregulation of expression or changes in enzyme activity in screening assays of compounds, for detection of any number of allelic variants of the gene, or for use as molecular markers in plant breeding programs. The isolated nucleic acids of the present invention can also be used for recombinant expression of their encoded polypeptides or for use as immunogens in the preparation and/or screening of antibodies. The proteins can be employed in assays for enzyme agonists or antagonists of enzyme function or for use of immunogens or antigens to obtain antibodies specifically immunoreactive with a protein.
Db	537 AGAMMALVRYSAVLTGPFITNLDDHYTNISKALREAGNEPLRVYVSRKPSOHKK	521	CC	
QY	582 DGDKDNDYARNNTVFDINRGLDGIQGVYVGTCVFNRTALYGEPEPKVKKKPSL	641	CC	
Db	477 EGGWMQDGTPFPGNNTXDHGMIOVFLGGNGDGAEGNLPRLVYVSRKPSOHKK	536	CC	
QY	522 AGAMMALVRYSAVLTGPFITNLDDHYTNISKALREAGNEPLRVYVSRKPSOHKK	581	CC	
Db	537 AGAMMALVRYSAVLTGPFITNLDDHYTNISKALREAGNEPLRVYVSRKPSOHKK	596	CC	
QY	582 DGDKDNDYARNNTVFDINRGLDGIQGVYVGTCVFNRTALYGEPEPKVKKKPSL	641	CC	
Db	597 DGDKDNDYARNNTVFDINRGLDGIQGVYVGTCVFNRTALYGEPEPKVKKKPSL	654	CC	

Also disclosed are a recombinant expression cassette comprising the polynucleotide (operably linked to a promoter) a host cell comprising the recombinant expression cassette and a transgenic plant comprising the recombinant expression cassette. The nucleic acid is useful in the improvement of stalk quality for improved stand or slake, and in the increased concentration of cellulose in the pericarp, hardening the kernel, and thus improving its handling ability. The nucleic acids may also be used as probes or amplification primers in the detection, quantification or isolation of gene transcripts, as probes in detecting deficiencies in the level of mRNA, for detecting gene mutations or allelic variants, for monitoring up regulation of expression or change in enzyme activity in screening assays, for site directed mutagenesis, and in sense or antisense suppression of one or more genes in a host cell, tissue or plant. The polypeptides may be used in assays for enzyme agonists or antagonists, and as immunogen or antigen to obtain antibodies specifically immunoreactive with the protein. The present sequence represents a corn cellulose synthase of the invention.

OY	642	LSKLGGSKKNSKKAKKEDKKSSGRHTSTVPPNLDDEEGVAGGFDDKEKKLMQ
Db	655	LSSLRGKSKRKKASKKGSDKKKSQKHVSPPVNELEDFEGVAGGFDDKEKKLMQ
OY	702	SLEKRGOSAVFASVTLMEENGVPSPATPENPLKTAHV-TSGYEDRSWDMEIGNTYG
Db	774	VTEDLITGPKMHARGWRSIYCMPKPKPAFKGSAPINLSDRLNQVLRWALGSVEILSRHCP
OY	714	SLEKRGOSAVFASVTLMEENGVPSPATPENPLKTAHV-TSGYEDRSWDMEIGNTYG
OY	822	IWGYNGRKLFLERRAYVTTIYTTSIPIALMCYCLLAVCLFTNQDPIQPQINSIWIFL
Db	834	LWPGGGRKFLERAYVTTIYTTSIPIALMCYCLLAVCLFTNQDPIQPQINSIWIFL
OY	882	SIFLISIFATGILEMRWVGVIDEMARNEOFWVIGGSVAHLFAVFGOILKVLAGIDNTV
Db	894	SIFLISIFATGILEMRWVGVIDEMARNEOFWVIGGSVAHLFAVFGOILKVLAGIDNTV
OY	942	TSKASDEDCDAELMFLKGMGRQNRTPTIVVVSVLLASIFSLIWWVRIDPFTSRVTGPDILE
Db	954	TSKASDEDCDAELMFLKGMGRQNRTPTIVVVSVLLASIFSLIWWVRIDPFTSRVTGPDILE
OY	1002	FATWWIVHLPFLKGLMGRQNRTPTIVVVSVLLASIFSLIWWVRIDPFTSRVTGPDILE
Db	1014	FATWWIVHLPFLKGLMGRQNRTPTIVVVSVLLASIFSLIWWVRIDPFTSRVTGPDILE
OY	1062	GINC 1065
Db	1074	GINC 1077
RESULT 8		
ID	AAU10496	AAU10496 standard; Protein; 1079 AA.
XX	XX	AC
XX	XX	AAU10496;
DT	14-FEB-2002	(first entry)
DE	Corn cellulose synthase Cqrail9/cesa-9.	
XX	KW	Corn cellulose synthase; Cggs15; cesa-3; Cqrail9; csaA-9;
KW	KW	stalk quality; improved stand; silage; pericarp; kernel hardening;
KW	KW	handling ability; transgenic plant; immunogen.
OS	zea mays.	
XX	PN	WO200179516-A2.
XX	PD	25-OCT-2001.
XX	PF	12-APR-2001; 2001WO-US11951.
PR	14-APR-2000; 2000US-0550483.	
XX	PA	(PION-) PIONEER HI-BRED INT INC.
PT	PT	Dhuga KS, Helentjaris TG;
XX	DR	WPI; 2002-0411338/05.
DR	DR	N-PSDB; AAS16458.
XX	PT	New cellulose synthase polypeptides and polynucleotides, useful in improving stalk quality or silage, and in increasing concentration of cellulose in the pericarp, hardening the kernel for improved handling ability -
PT	XX	Claim 3; Page 85-87; 88pp; English.
PS	XX	The invention relates to isolated nucleic acids encoding two cellulose synthase proteins from corn, Cggs15 (cesa-3) and Cqrail9 (cesa-9).
CC	CC	
CC	CC	

Also disclosed are a recombinant expression cassette comprising the poly nucleotide (operably linked to a promoter) a host cell comprising the recombinant expression cassette and a transgenic plant comprising the recombinant expression cassette. The nucleic acid is useful in the improvement of stalk quality for improved stand or silage, and in the increased concentration of cellulose in the pericarp, hardening the kernel, and thus improving its handling ability. The nucleic acids may also be used as probes or amplification primers in the detection, quantification or isolation of gene transcripts, as probes in detecting deficiencies in the level of mRNA, for detecting gene mutations or allelic variants, for monitoring up regulation of expression or changes in enzyme activity in screening assays, for site directed mutagenesis, and in sense or antisense suppression of one or more genes in a host cell, tissue or plant. The polypeptides may be used in assays for enzyme agonists or antagonists, and as immunogen or antigen to obtain antibodies specifically immunoreactive with the protein. The present sequence represents a corn cellulose synthase of the invention.

OY	701	MSLEKKRGQSAAVPAVSTLIMENGYPSPSATPENILKEIHWISCGEDKSDWGMELCWIG	760
Db	715	MSLEKKRGQSAAVPAVSTLIMENGYPSPSATPENILKEIHWISCGEDKSDWGMELCWIG	774
QY	761	SVTEILDGFPMHARGWRISYCMRPLPATRGSAPINLSDRLNQVRWALGSVEILFSRH	820
Db	775	SVTEILDGFPMHARGWRISYCMRPLPATRGSAPINLSDRLNQVRWALGSVEILFSRH	834
OY	821	PWYGYNGLKEFRAVNTTIPITSIPLMVCYLAVCLFTNFIPIQNSNTASIW	880
QY	835	PLWYGYYGRKLKFLEFRAYINTTIVPLSLPLLYCILPAICLTGKPEIPELSNFAWI	894
Db	881	LSIFLSIFATGLEMWRWSSVGIDDMWRNQFWVIGVSAHLFAVYQGIKVLLAGIDNTF	940
Db	895	ISLFISIFATGLEMWRWSSVGIDDMWRNQFWVIGVSAHLFAVYQGIKVLLAGIDNTF	954
OY	941	VTKASDEDGDFEFLYLKRWTLLIPPTILLVNVNGVAGVSYAINTSGYOSWGLPLGR	100
QY	955	VTKASDEDGDFEFLYLKRWTLLIPPTILLVNVNGVAGVSYAINTSGYOSWGLPLGR	101
Db	1001	FFAWFWVTVHLYPLFKLGMQRQNRTPTIVVWSVLAISFLWVDRPFTSRVTPDILE	106
Db	1015	FFAWFWVTVHLYPLFKLGMQRQNRTPTIVVWAILLASFLWVDRPFTSRVTPDILE	107
OY	1061	CGINC 1065	
Db	1075	CGINC 1079	
RESULT 9			
ID	AY84121	AY84121 standard; Protein: 1076 AA.	
XX	XX		
AC	AY84121;		
XX	XX		
DT	03-JUL-2000	(first entry)	
XX	XX		
DE	DNA	encoding a maize cellulose synthase.	
KW	Maize; cellulose synthase; stalk quality; stand; silage; cellulose; transgenic plant; plant breeding marker.		
KW	zea mays.		
OS	XX		
FH	Key	Location/Qualifiers	
FT	Misc-difference	425	
FT	/note=	"encoded by ANG"	
PN	WO200009706-A2.		
XX	PD	24-FEB-2000.	
XX	PF	16-AUG-1999; 99WO-US18760.	
PR	17-AUG-1998;	98US-0096822.	
XX	PA	(PION-) PIONEER HI-BRED INT INC.	
XX	PI	Dhuga KS, Helentjaris TG, Bowen BA, Wang X;	
XX	DR	WPI; 2000-224343/19.	
XX	DR	N-PSDB; AAZ299533.	
Claim 15; Page 196-199; 119pp; English.			
CC	The present sequence represents a maize cellulose synthase polypeptide.		
CC	The cellulose synthase can be used for the improvement of stalk quality for improved stand or silage. It also provides an increased concentration of cellulose in the pericarp, hardening the kernel and improving its		
CC	New genes which encode maize cellulose synthase polypeptides in plants useful for modulating the expression of cellulose synthase in plants and to produce transgenic plants expressing the novel protein		
XX	Claim 15; Page 196-199; 119pp; English.		

CC handling ability. The sequences are used to produce transgenic plants CC and seeds expressing the cellulose synthase. The polynucleotide is CC used for modulating, preferably increasing, the level of the synthase CC in a plant cell. The plants are preferably monocots. The polynucleotide CC is also used as a probe or primer in the detection, quantitation or CC isolation of gene transcripts. The probes are useful in detecting CC deficiencies in the level of mRNA in screenings for desired transgenic CC plant, for detecting mutations in the gene, for monitoring upregulation CC of expression or changes in enzyme activity in screening assays of CC compounds, for detection of any number of allelic variants of the gene, CC or for use as molecular markers in plant breeding programs. The CC isolated nucleic acids of the present invention can also be used for recombinant expression of their encoded polypeptides or for use as CC immunogens in the preparation and/or screening of antibodies. The CC proteins can be employed in assays for enzyme agonists or antagonists, CC or enzyme function or for use of immunogens or antisera to obtain CC antibodies specifically immunoreactive with a protein.

Db 714 SLIKRFGGSAAFVASTIMYEYGPQSAATPESULIKEAHTHVISCGYEDKTEWGEIGWIGS 773
 QY 762 VTEDILTGKFMHARGWRSIYCMLPKPFGSAPINSLSDRLNQYLRWALGSVEILFSRHC P 821
 CC ||||||| ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
 CC ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
 Db 774 VTEDILTGKFMHARGWRSIYCMLPKPFGSAPINSLSDRLNQYLRWALGSVEILFSRHC P 833
 QY 822 IWTGYNGRKLFLERFAVNTTIPITSPLLMCITLAVCLFTNQFIPQISIASTWV 881
 CC :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 CC :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 CC 834 LWYGYYGRKLFLERFAVNTTIPITSPLLMCITLAVCLFTNQFIPQISIASTWV 893
 Db 882 SLELSIFFTGLEMRWSGVGIDEMWRNEQFWIGVSVASHLFAVFOGLIKVLAGIDTNFTV 941
 QY ||||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 CC 894 SLELSIFFTGLEMRWSGVGIDEMWRNEQFWIGVSVASHLFAVFOGLIKVLAGIDTNFTV 953
 Db 942 TSKASDEDGDFAEYLFWKTTLIPPTMLTIVLNUVGVAVGVSAINSYQSNQPLEGKF 1001
 QY ||||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 Db 954 TSKASDEDGDFAEYLFWKTTLIPPTMLTILINLVGVVAGISYAINSQSMQPLFGKLF 1013
 QY 1002 FAWFVVIVLWYPEFKLGKGRQNRPPTIVWW 1031
 Db 1014 FAFWVIVLWYPEFKLGKGRQNRPPTIVWW 1043

RESULT 11

ID AAY84109
 ID AAY84109 standard; Protein: 1075 AA.
 XX AC AAY84109;
 XX DT 03-JUL-2000 (first entry)
 DE Amino acid sequence of a maize cellulose synthase.
 KW Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
 XX OS Zea mays.
 XX Key Location/qualifiers
 FT Misc-difference 245 /note= "encoded by NAT"
 XX wo200009706-A2.
 PD 24-FEB-2000.
 XX PF 16-AUG-1999; 99WO-US18760.
 XX PR 17-AUG-1998; 98US-0096822.
 XX PA (PION-) PIONEER HI-BRED INT INC.
 XX PT Dhuogga KS, Helentjaris TG, Bowen BA, Wang X;
 XX DR WII; 2000-224343/19.
 DR N-PSDB; AAZ9349/7.
 XX PS Claim 15; Page 102-105; 119pp; English.

New genes which encode maize cellulose synthase polypeptides in plants useful for modulating the expression of cellulose synthase in plants and to produce transgenic plants expressing the novel protein

XX

The present sequence represents a maize cellulose synthase polypeptide. The cellulose synthase can be used for the improvement of stalk quality for improved stand or silage. It also provides an increased concentration of cellulose in the pericarp, hardening the kernel and improving its handling ability. The sequences are used to produce transgenic plants and seeds expressing the cellulose synthase. The polynucleotide is used for modulating, preferably increasing, the level of the synthase in a plant cell. The plants are preferably monocots. The polynucleotide is also used as a probe or primer in the detection quantitation or isolation of gene transcripts. The probes are useful in detecting

CC deficiencies in the level of mRNA in screenings for desired transgenic plant, for detecting mutations in the gene, for monitoring upregulation of expression or changes in enzyme activity in screening assays of compounds, for detection of any number of allelic variants of the gene, or for use as molecular markers in plant breeding programs. The isolated nucleic acids of the present invention can also be used for recombinant expression of their encoded polypeptides or for use as immunogens in the preparation and/or screening of antibodies. The proteins can be employed in assays for enzyme agonists or antagonists of enzyme function or for use of immunogens or antigens to obtain antibodies specifically immunoreactive with a protein.

XX SQ Sequence 1075 AA;

Query Match 71.0%; Score 4028.5; DB 21; Length 1075;
 Best Local Similarity 69.6%; Pred. No. 0;
 Matches 752; Conservative 130; Mismatches 153; Indels 45; Gaps 12;

QY 4 EGSTAG--KPMKNIIVPQPCQCSDNVKTVTDREVAACDQSFPPVRCYEVKRGQNS 61
 23 DGDVPGSAKPTKSANGQVQCICGDSVQGVSATGDVFVACNECAFPVCRPCYEVKGNQC 82

Db 62 CPOCKTRYKRKLGSPAPGKDDEGDAEGTVEFNYPQEKEISERLGMWHLTGKCEEM 121
 Db 83 CPCCKTRYKRKLGSPAPGKDDEGDAEGTVEFNYPQEKEISERLGMWHLTGKCEEM 128
 QY 122 EPOYDKEVS-----HNHLPRLTSRDTSGERSAASPERVSSTTAGGKRLPYSSDVN 174
 Db 129 LOGDDADLSSARHEPHRIRPRLTSQOISGEPDASPDHSI-----RSPTSYVD 180

QY 175 QS--PNRSLVDP----VGIGNVAKRERDWMKQEKNGTPVSTOAASERGGVVIDAS 226
 181 PSVPVPVRLVDPKSDLNQYGLMSVDRKVERESWRVQDKDNNMQVTNKPYPEARGG-DME-G 238

Db 227 TDILADEALLNDEEARPLSRKKSIPPSSRINPRMIMRLVILCLHYRINTPVNAFA 286
 Db 239 TGSNGEXWQMVDDARPLSLRIVPISSNQNLNLYRVVITLRLTCFFPQYRSHPVDAYG 298
 QY 287 LWLVSVICEIWPLAISWLDDQFPKWPVPPNRETYLDRAILYDREGESQSLAVIDIFSTVD 346
 Db 299 LWLVSVICEIWPLAISWLDDQFPKWPVNPRTYLDRAILYDREGESQSLAVIDIFSTVD 358

QY 347 PIKEPPLMTANTVLSILAVDYPDKVSCYVFDGAMLSFSLAESEFARKWNPCKRY 406
 Db 359 PIKEPPLMTANTVLSILAVDYPDKVSCYVFDGAMLSFSLAESEFARKWNPCKKH 418

QY 407 SIEPRAPAEWYFAAKIDLKDKYQTSVYKDRRAMKREVEEKFIRINALVSKALKCPBGWV 466
 Db 419 NTEPRAPAEWYFAOKTDYLDKIQPSFVKERRAKREVEEKFIRINALVSKALKCPBGWV 478

QY 467 MDGTPWPWNQNTCDHPGMQIVLQLQNGDLAECNEILPRLVYVSRKRGFQHHKKAGMN 526
 Db 479 MADGTAPGNPNFRDHPRMIOVFGHSGGLDTPDGNEUPRLVYVSRKRGFQHHKKAGMN 538

QY 527 ALVRVSAVLTNGFILNLCDAVINNARLAMCFJMDPNIGKQCYVOPORFIDK 586
 Db 539 ALRVSATLNGAYLNLCDAVINNARLAMCFJMDPNIGKQCYVOPORFIDK 598

PS 587 NDQYANRNTVFDINLNLQDGTQGPVWVGTGEGVNFNTALYCEPPTRVKHKRPSLASKL 646

Db 599 HDTYANRNTVFDINLNLQDGTQGPVWVGTGEGVNFNTALYCEPPTRVKHKRPSLASKL 658

QY 647 GSRKRKNKAKKESDKKKSGRITDSTPVFDNDDIEGVEGAGFDEKALLMSQMLEK 706
 Db 659 CGRRKRNKSYMDRSRIMR-TESSAIPNMEDIEBEGE---GYEDERSVLSRQKLER 715

QY 707 FGSRVFASTIMENGGPPSAPTAPEINLKEAHVHSYEDKSDWMEIGYGSTED 766
 Db 716 FGQSPFIASTPMTQGGPSTPNSLKEAHVHSYEDKSDWMEIGYGSTED 775

QY 767 LIGFKMHARGWRSIYCMLPKPFGSAPINSLSDRLNQYLRWALGSVEILFSRHCPTWGY 826
 Db 776 LIGFKMHARGWRSIYCMLPKPFGSAPINSLSDRLNQYLRWALGSVEILFSRHCPTWGY 835

Db 956 DEGDFEALYFLVFRWTSLLIPPTVLVNLGVAGISAYNSGYQSGWPLRGKLFSSIW 1015
 Qy 1007 IVHLYPFLKGLMGQRNRTPTIVWWSLLASIFSLWWRIDPFTSRVTPDIL-EGCINC 1065
 Db 1016 IHLYPFLKGLMGQRNRTPTIVWWSLLASIFSLWVKIDFISPDKAAALGQCGVNC 1075
 RESULT 14
 AAY84110
 ID AAY84110 standard; Protein: 1074 AA.
 AC AAY84110;
 XX DT 03-JUL-2000 (first entry)
 DE Amino acid sequence of a maize cellulose synthase.
 KW Maize; cellulose synthase; stalk quality; stand; silage; cellulose;
 KW transgenic plant; plant breeding marker.
 OS zea mays.
 XX PN WO200009706-A2.
 XX PD 24-FEB-2000.
 XX PF 16-AUG-1999; 99WO-US18760.
 PR 17-AUG-1998; 98US5-0096822.
 PA (PION-) PIONEER HI-BRED INT INC.
 XX PT Dhugga K.S., Helentjaris T.C., Bowen B.A., Wang X.;
 XX DR WPI; 2000-224343/19.
 DR N-PSDB; AAZ299500.
 XX PT New genes which encode maize cellulose synthase polypeptides in plants useful for modulating the expression of cellulose synthase in plants and to produce transgenic plants expressing the novel protein -
 XX PS Claim 15; Page 110-113; 119pp; English.
 XX The present sequence represents a maize cellulose synthase polypeptide. The cellulose synthase can be used for the improvement of stalk quality for improved stand or silage. It also provides an increased concentration of cellulose in the pericarp, hardening the kernel and improving its handling ability. The sequences are used to produce transgenic plants used for modulating, preferably increasing, the level of the synthase in a plant cell. The plants are preferably monocots. The polynucleotide is also used as a probe or primer in the detection, quantitation or isolation of gene transcripts. The probes are useful in detecting deficiencies in the level of mRNA in screenings for desired transgenic plant, for detecting mutations in the gene, for monitoring upregulation of expression or changes in enzyme activity in screening assays of the gene, compounds, for detection of any number of allelic variants of the gene, or for use as molecular markers in plant breeding programs. The isolated nucleic acids of the present invention can also be used for recombinant expression of their encoded polypeptides or for use as immunogens in the preparation and/or screening of antibodies. The proteins can be employed in assays for enzyme agonists or antagonists of enzyme function or for use of immunogens or antigens to obtain antibodies specifically immunoreactive with a protein.
 SQ Sequence 1074 AA;

Query Match 70.9%; Score 4024; DB 21; Length 1074;
 Best Local Similarity 69.4%; Pred. No. 0;
 Matches 749; Conservative 128; Mismatches 159; Indels 44; Gaps 10;

Db 23 DGDAPVPAKPTKSANGQVQCQICGDTVGVSATGDFWVACNECAFPCRCPEYBRKEGNQC 82
 Qy 62 CPOCKTRKRLKGSPALPGKDDEGLADEGTVEFFNPKOKI1SERMLGWHLTRKGGEEMG 121
 Db 83 CPOCKTRKRLKGSPALPGKDDEGLADEGTVEFFNPKOKI1SERMLGWHLTRKGGEEMG 128
 Qy 122 EPOYDKEVS-----HNHLPLRISRODTSGERSAASPERLSVSTIAGKRLPQYSDVN 174
 Db 129 LQGDDADLSSSARHDHPHRIPRLISQGQISGEPPDASDRHSI-----RSPTSSVD 180
 Qy 175 QS--PNRRIYDP-----VGLGMWAMERVDGKMKDEKNTEPVSIAASERGGVIDAS 226
 Db 181 PSVFPVPVRIVDPSKDLNSYGLNSWDKERVESWRVSKVQDKNMQLQTNKYPEARGDME--G 237
 DE 227 TDIADEALLNDBARQPLRSRKVISPPSRINPRMNRIMRLVILCFLHIRTNPVNAFA 286
 Db 238 TGSNGEDKOMVDAARLPLSRTVPLISSNOLNLYRIRPDRGKPLQFQYRSPVNAFA 297
 Qy 287 LWLWSVCEIWFAISWLQDFPKWPKPWFVNPVRETYLDLALRYDREGEPSQALAVDIFVSTD 346
 Db 298 LWLWSVCEIWFAISWLQDFPKWPKPWFVNPVRETYLDLALRYDREGEPSQALAVDIFVSTD 357
 Qy 347 PIKEPPTVANTVLSIADVDPVTKVSYVEDGAAMLSFEIASAEFARKWWPEKKY 406
 Db 358 PLKEPLPLITANTVLSIADVDPVTKVSYVEDGAAMLSFEIASAEFARKWWPEKKY 417
 Qy 407 SIEPPAPEWFAAKIDYKDKQTSFKDVRKRMREBEFKRKLINALVSKALKCPEEW 466
 Db 418 NIERPAPPEYEAKIDYKDKQTSFKDVRKRMREBEFKRKLINALVSKALKCPEEW 477
 Qy 467 MDGQPWGRNNNTCDHPGMQIQLVFLGQNGLDAEGNEELPRLVY'SREKRGFOHHKAGMN 526
 Db 478 MADCTAWFGNNPRDHPGMQLQEVGHSGGLDTONELPRLVY'SREKRGFOHHKAGMN 537
 Qy 527 ALVRSVATLNGSFILNIDCRVTTNSALREMCFLANDPNIGKQCVVOETPFDGIDK 586
 Db 538 NDIVYANRWFEDINLRGLDGTOGPVYVTGCVFNRATLGCEPPIKVHKRPSLKL 646
 Qy 587 NDIVYANRWFEDINLRGLDGTOGPVYVTGCVFNRATLGCEPPIKVHKRPSLKL 597
 Db 598 HDIYVANRWFEDINLRGLDGTOGPVYVTGCVFNRATLGCEPPIKVHKRPSLKL 597
 Qy 647 GGSRKKNKAKKEDKKKSGRHTDSTVVFNUFLDIEEGVEAGPDDDEKALLMSQMSLEKR 706
 Db 658 CGRRKRKNSYMDQSRSRMR-TESSAIPNFMEDIEGIE-GYEDERSVLSQKLER 714
 CC 707 FGGSAAVEFASTLMEENGYPSPATPENLIKATIVVSCYEDKSDWGMIGWIGVSVEDI 766
 CC 715 FGQSPIFIATSTFMTOGGIPSPSTNPASLKEATHWISCGIEDKTEWKGREGWIGVSVEDI 774
 CC 767 LIGFKMHARGWRSIYCMLPKLPAFKGSAPINLSPRNQVLWALGSEVELFSHCPINGY 826
 CC 775 LIGFKMHARGWRSIYCMLPKLPAFKGSAPINLSPRNQVLWALGSEVELFSHCPINGY 834
 Qy 827 NGRLKFLERFAYVNTTIPITSPLLMCTLLAWCLFNQFIIPQISIASTWFLSLS 886
 Db 835 NGRKLLRLAYINTIVPITSPLIPLAVCPLACIUNKTFPEISYVAGHEFILFAS 894
 CC 887 IFAMGILEMRWMSGVIDEMWRNIFQFWVIGVS AHLFAVFOGLIKVLQAGIDTNFTVSKAS 946
 CC 895 IFATGILEMRWMSGVIDEMWRNIFQFWVIGVS AHLFAVFOGLIKVLQAGIDTNFTVSKAS 954
 Qy 947 DEGDFEALYFLWKTTLIPPTPLTLLVLUVGVAGVSTAISNGYOSNGPLGFLFEEFW 1006
 Db 955 DEGDFEALYFLWKTTLIPPTPLTLLVLUVGVAGVSTAISNGYOSNGPLGFLFEEFW 1014
 Qy 1007 IVHLYPFLKGLMGQRNRTPTIVWWSLLASIFSLWWRIDPFTSRVTPDIL-EGCINC 1065
 Db 1015 IHLYPFLKGLMGQRNRTPTIVWWSLLASIFSLWVKIDFISPDKAAALGQCGVNC 1074
 RESULT 15

AAV84112	Db	129 LQGDDADLSSARAHDPHIRIPRTSGQOISGPASPDHSI-----RSPTSSYVD 180
ID ANY84112 standard; Protein: 1074 AA.	Id	175 OS--PNRRTVDP----VGLGNAWAKERDVGKMKRKDKNTGPVSTQASERGVVIDAS 226
XX AC AAV84112;	Qy	181 PSVPVPVRIVDPSKDLNSYGLNSVWDWKERESWRVKQDKRNMLQVTKYPEARGDME--G 237
XX DT 03-JUL-2000 (first entry)	Db	227 TDILADEALLNDEARRQPLRSKRVSIPSSRSPSRINPRTGPMVIRLWLTCLEHYRTNPVPAFA 286
XX DE Amino acid sequence of a maize cellulose synthase.	Qy	238 TGSNGEDMOMVNDARRPLSRTIVPSSQNOLNRYTIVTILICFFOYRISHPVRAYG 297
XX KW Maize; cellulose synthase; stalk quality; stand; silage; cellulose;	Db	287 LWLVSVICIWALSWILDQOPKWFVWRETYLDRLAIRYDRGEPSOLAANDIFNSTVD 346
XX KW transgenic plant; plant breeding marker.	Qy	OS Zea mays.
OS Zea mays.	Db	298 LWLVSVICIWALSWILDQOPKWFVWRETYLDRLAIRYDRGEPSOLAANDIFNSTVD 357
XX PN WQ200009706-A2.	Qy	XX
XX PR 17-AUG-1998; 98US-0096822.	Db	347 PLKEPPVLTANTVLSLAVDYPVDKVSVICVSDGSAMLTFESISETREFARKWVPCKKH 417
XX PA (PION-) PIONEER HI-BRED INT INC.	Db	358 PLKEPPVLTANTVLSLAVDYPVDKVSVICVSDGSAMLTFESISETREFARKWVPCKKH 417
XX PT Dhuuga KS, Helentjaris TG, Bowen BA, Wang X;	Qy	XX
XX DR WPI: 2000-224343/19.	Db	407 SIEPRAPWYFAAKIDYLKVQTSFVYDQRKREYEEFKTRINALVSKALKCPEPGW 466
XX N-PSDB; AAZ99506.	Qy	418 NIEPRAPEFYFAOKIDYLKDTQPSFVCKRERRAKREYEEFKTRINALVAKAOVPEBKG 477
PS Claim 15; Page 126-128; 119pp; English.	Db	467 MDGTPWIGGNNTGHPGMIQVFLGQNGGLDAEGNEELRLVYVSRKRGFOHHKAGMN 526
PT New genes which encode maize cellulose synthase polypeptides in plants useful for modulating the expression of cellulose synthase in plants and to produce transgenic plants expressing the novel protein	Qy	478 MAQTAIWGGNNPRDHPGMIQVFLGHSQGLDTQNEELRLVYVSRKRGFOHHKAGMN 537
PT PT useful for modulating the expression of cellulose synthase in plants and to produce transgenic plants expressing the novel protein	Qy	527 ALYRVSAYLTINGPFLNLDCDHYTNNSKALREAMCFLMDPNTGKQCVYQFORDGIDK 586
XX C-claim 15; Page 126-128; 119pp; English.	Db	538 ALTVRSAVLTINGAYLLNLVCDHYNFNSKALREAMCFMDPALGRKTCVYQFORDGIDL 597
CC The present sequence represents a maize cellulose synthase polypeptide. The cellulose synthase can be used for the improvement of stalk quality for improved stand or silage. It also provides an increased concentration of cellulose in the pericarp, hardening the kernel and improving its handling ability. The sequences are used to produce transgenic plants and seeds expressing the cellulose synthase. The polynucleotide is used for modulating, preferably increasing, the level of the synthase in a plant cell. The plants are preferably monocots. The polynucleotide is also used as a probe or primer in the detection quantitation or isolation of gene transcripts. The probes are useful in detecting deficiencies in the level of mRNA in screenings for desired transgenic plant, for detecting mutations in the gene, for monitoring upregulation of expression or changes in enzyme activity in screening assays of compounds, for detection of any number of allelic variants of the gene, or for use as molecular markers in plant breeding programs. The isolated nucleic acids of the present invention can also be used for recombinant expression of their encoded polypeptides or for use as immunogens in the preparation and/or screening of antibodies. The proteins can be employed in assays for enzyme agonists or antagonists of enzyme function or for use of immunogens or antigens to obtain antibodies specifically immunoreactive with a protein.	Qy	587 NDRYANRNTWVFEDINRGLDGTDQGPPVYVTGCVENRPLAEGEPPIVKHKKPEPSLUSKL 646
CC XX	Db	598 HDRYANRNLVFFDLMKGLDLOGPVVYGTGCCFNROALYGDPLVTEADLPNVVHSC 657
CC PS Claim 15; Page 126-128; 119pp; English.	Qy	647 GGSRKKNKAKKESDKKSGRHDSTVPUFNLDIDESVEGAGFDEKALMSOMSLER 706
CC CC	Db	658 CGRRKRKNKSYMDQSRRIMR-TESAPATFNMEDIEBIE--GYEDERSVLMSSORKLEKR 714
CC CC	Qy	707 FGOSAVFVASTLMEENGGUPPSATPENLKEAHLVSCGIEDKSDWMGMRIGWIGSYTEDI 766
CC CC	Db	715 FGQSPPIFLASTEMQGPGTSPNPAKSAHVAHVICGEGDTEWGREGIWIGSYTEDI 774
CC CC	Qy	767 LTERPKMHARGWRSYCMRKPARKGASDINSLRNLQYRLWALGSEVELFSRCPWGY 826
CC CC	Db	775 LTGFKFMHARGWQSIYCMPPRCFKGKSAPINLSDRNLQYRLWALGSEVELLSRCPWGY 834
CC CC	Qy	827 NGSLKFLRFAVNTTIPITSPPLMYCTLAVALCFTNOFTIPOSNIASTWELSFPLS 886
CC CC	Db	835 NGRLKLLERLAYINTIVPITSVPLTACVCLPACIQLNKFLPETISNYAGMFILLFAS 894
CC CC	Qy	887 IFATGILEMRWSSVGIDEWWRNQFWVYGGVSAHLFLAVFOGLIKVLAGIDINFVWSKAS 946
CC CC	Db	895 IFATGILELRWSVGIEDWWRNQFWVIGTSAAHLFLAVFOGLIKVLAGIDINFVWSKAS 954
CC CC	Qy	947 DEDGDAEFLYLFKWTTLIAPPTRILVNWVAGVSTAINSGYQSGPLFKEFREWV 1006
CC CC	Db	955 DEGDGAELYLFKWTTLIAPPTRILVNWVAGVSTAINSGYQSGPLFKEFREWV 1014
SQ Sequence 1074 AA;	Qy	1007 IVILXPFLKGMLGRQRNRUPTIVWVWVSLASIFSLAWRIDPFTSRTGPDTL-EGINC 1065
Query Match 70.9%; Score 4024; DB 21; Length 1074; Best Local Similarity 69.4%; Pred. No. 0; Matches 749; Conservative 128; Mismatches 159; Indels 44; Gaps 10;	Db	1015 IHLHPLKLGMLGRQRNRUPTIVWVWVSLASIFSLAWRIDPFTSRTGPDTL-EGINC 1074
Qy 2 ESRGETACKPMKNAIVPTCQICSDNVNGKTVNDGDRFVACDICSPVCRCYEVERKDQNS 61	Search completed: June 16, 2003, 11:06:41	
Db 23 DGDAPVPAKPTKSANGQVQICGDTIVGVSATGDFVFAACNECAFPCRCYEVERKEGNQC 82	Job time : 50 secs	
Qy 62 CPOCKTRYKRLKGSPALPGDKDEGLADEGTVFENYPOKEKTSERMLGWHLTRKGEMG 121		
Db 83 CPOCKTRYKROKGSPRVHGDEBEDV-DDLDNEFYKQ-----GNGKPEWQ 128		
Qy 122 EPGDKEWS-----INHLPRLTSRQDTSGEFASASPERLVSSTAGKALPYSSDW 174		

Title:	AAC39336	34	103	1.8	470	1	ROCC_BAGSU	P39636 bacillus su
Perfect score:	5677	35	103	1.8	1066	1	NUC2_NEUCR	P01317 neurospora
Sequence:	1 MESEGETAGKPMKNIVPQTC.....IDPFITSRVTPGPDIILECGINC 1065	36	103	1.8	1232	1	B3A3_HUMAN	P48751 homo sapien
Run on:	June 16, 2003, 11:03:13 ; Search time 15 Seconds	37	102	1.8	642	1	ARE2_YEAST	P53629 saccharomyces
	{without alignments}	38	101.5	1.8	1863	1	BRC1_HUMAN	P28398 homo sapien
	(2944.819 Million cell updates/sec	39	101	1.8	617	1	VAA1_BOVIN	P31404 bos tauris
		40	101	1.8	3097	1	CADN_DROME	Q15943 drosophila
		41	100.5	1.8	744	1	YTV2_CASER	Q18600 caenorhabditis
		42	100.5	1.8	1020	1	ATCL_DROME	P22700 drosophila
		43	99.5	1.8	322	1	NUIM_STREU	P15548 strongyloides
		44	99.5	1.8	536	1	COX1_ALIMA	P80440 allomyces m
		45	99.5	1.8	714	1	HUNK_MOUSE	Q88866 mus musculus
Searched:	112892 seqs, 41476328 residues							
Total number of hits satisfying chosen parameters:	112892							
Minimum DB seq length:	0							
Maximum DB seq length:	200000000							
Post-processing:	Minimum Match 0 %							
	Maximum Match 100 %							
	Listing first 45 summaries							
Database :	SwissProt_40:*							
Pred.	No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.							
	SUMMARIES							
Result No.	Score	Query Match Length	DB ID	Description				
1	273	4.8	872	1 BCSA_ECO57	RESULT 1			
2	273	4.8	874	1 BCSA_SALTI	BCSA_ECO57	STANDARD;	PRT:	872 AA.
3	272	4.8	872	1 BCSA_ECOLI	ID: BCSA_ECO57			
4	271	4.8	874	1 BCSA_XANAC	AC: 08X5L7;			
5	252.5	4.4	729	1 ACS2_ACEXY	DT: 15-JUN-2002 (Rel. 41, Last sequence update)			
6	252.5	4.4	1596	1 ACS2_ACEXY	DT: 15-JUN-2002 (Rel. 41, Last annotation update)			
7	247.5	4.4	1518	1 BCA4_ACEXY	DT: 15-JUN-2002 (Rel. 41, Last annotation update)			
8	247.5	4.4	1518	1 BCA5_ACEXY	Cellulose Syntthase catalytic subunit [lapp-forming] (EC 2.4.1.12).			
9	236.5	4.2	756	1 BCA2_ACEXY	DE: BCSA OR 24948 OR ECS4113.			
10	228.5	4.0	754	1 BCA1_ACEXY	GN: Escherichia coli O157:H7;			
11	228	4.0	739	1 BCSA_PSEFL	OC: Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
12	208.5	3.7	745	1 BCA3_ACEXY	OX: Escherichia coli O157:H7;			
13	196	3.5	1550	1 AGS1_ACEXY				
14	119	2.1	1 COX1_PARLI	1 P21877 acetobacter	[1]			
15	118.5	2.1	590	1 YKPA_CABEL	SEQUENCE FROM N.A.			
16	115	2.0	357	1 GOFT_HUMAN	SEQUENCE FROM N.A.			
17	115	2.0	517	1 COX1_ASPIR	STRAIN=O157:H7 / RIMD 050952;			
18	110.5	1.9	441	1 YKDO_ECOLI	MEDLINE-21156231; PubMed-1258796;			
19	110.5	1.9	672	1 A1B_HUMAN	MEDLINE-21156231; PubMed-1258796;			
20	110	1.9	517	1 COX1_SRPPU	RAYASHI_T., MAKINO_K., OHNISHI_M., OGAWARA_N., YASUNAGA_T., RAHMAN_A., SHIBA_T., HATTORI_M., SHINAGAWA_H.;			
21	109	1.9	830	1 FARL_YEARST	RIDA_T., TAKAMI_H., HONDA_T., SASAKAWA_C., KUROKAWA_K., ISHLI_K., YOKOYAMA_K., RAHMAN_A., SHIBA_T., HATTORI_M., SHINAGAWA_H.;			
22	108	1.9	517	1 COX1_PISOC	P19449 acetobacter			
23	108	1.9	1029	1 RLP3_RAT	P58931 pseudomonas			
24	107.5	1.9	1437	1 MRP5_HUMAN	09wx61 acetobacter			
25	107	1.9	3033	1 POLG_HOVJ6	P21877 acetobacter			
26	106	1.9	357	1 GSP7_CANFEA	P12700 paracentrotus			
27	105.5	1.9	1076	1 Y338_SCIRPO	P34322 caenorhabditis			
28	105.5	1.9	4829	1 BTR6_HUMAN	P35575 homo sapien			
29	104.5	1.8	309	1 YESP_BACSU	Q33820 asterina pe			
30	104	1.8	995	1 ARH7_ARATH	P75905 escherichia			
31	104	1.8	1167	1 CLAA_BACTU	Q9y293 homo sapien			
32	104	1.8	1879	1 POIN_SMW1	P15544 strongyloides			
33	103	1.8			P21268 saccharomyces			
					P25001 pisaster oc			
					Ogere6 rattus norvegicus			
					O15440 homo sapien			
					P26660 hominis po			
					O19133 canis familiaris			
					P13892 schizosaccharomyces			
					P25001 pisaster oc			
					Ogere6 rattus norvegicus			
					O15440 homo sapien			
					P26660 hominis po			
					O19133 canis familiaris			
					P13892 schizosaccharomyces			
					P25001 pisaster oc			
					Ogere6 rattus norvegicus			
					O15440 homo sapien			
					P26660 hominis po			
					O19133 canis familiaris			
					P13892 schizosaccharomyces			
					P25001 pisaster oc			
					Ogere6 rattus norvegicus			
					O15440 homo sapien			
					P26660 hominis po			
					O19133 canis familiaris			
					P13892 schizosaccharomyces			
					P25001 pisaster oc			
					Ogere6 rattus norvegicus			
					O15440 homo sapien			
					P26660 hominis po			
					O19133 canis familiaris			
					P13892 schizosaccharomyces			
					P25001 pisaster oc			
					Ogere6 rattus norvegicus			
					O15440 homo sapien			
					P26660 hominis po			
					O19133 canis familiaris			
					P13892 schizosaccharomyces			
					P25001 pisaster oc			
					Ogere6 rattus norvegicus			
					O15440 homo sapien			
					P26660 hominis po			
					O19133 canis familiaris			
					P13892 schizosaccharomyces			
					P25001 pisaster oc			
					Ogere6 rattus norvegicus			
					O15440 homo sapien			
					P26660 hominis po			
					O19133 canis familiaris			
					P13892 schizosaccharomyces			
					P25001 pisaster oc			
					Ogere6 rattus norvegicus			
					O15440 homo sapien			
					P26660 hominis po			
					O19133 canis familiaris			
					P13892 schizosaccharomyces			
					P25001 pisaster oc			
					Ogere6 rattus norvegicus			
					O15440 homo sapien			
					P26660 hominis po			
					O19133 canis familiaris			
					P13892 schizosaccharomyces			
					P25001 pisaster oc			
					Ogere6 rattus norvegicus			
					O15440 homo sapien			
					P26660 hominis po			
					O19133 canis familiaris			
					P13892 schizosaccharomyces			
					P25001 pisaster oc			
					Ogere6 rattus norvegicus			
					O15440 homo sapien			
					P26660 hominis po			
					O19133 canis familiaris			
					P13892 schizosaccharomyces			
					P25001 pisaster oc			
					Ogere6 rattus norvegicus			
					O15440 homo sapien			
					P26660 hominis po			
					O19133 canis familiaris			
					P13892 schizosaccharomyces			
					P25001 pisaster oc			
					Ogere6 rattus norvegicus			
					O15440 homo sapien			
					P26660 hominis po			
					O19133 canis familiaris			
					P13892 schizosaccharomyces			
					P25001 pisaster oc			
					Ogere6 rattus norvegicus			
					O15440 homo sapien			
					P26660 hominis po			
					O19133 canis familiaris			
					P13892 schizosaccharomyces			
					P25001 pisaster oc			
					Ogere6 rattus norvegicus			
					O15440 homo sapien			
					P26660 hominis po			
					O19133 canis familiaris			
					P13892 schizosaccharomyces			
					P25001 pisaster oc			
					Ogere6 rattus norvegicus			
					O15440 homo sapien			
					P26660 hominis po			
					O19133 canis familiaris			
					P13892 schizosaccharomyces			
					P25001 pisaster oc			
					Ogere6 rattus norvegicus			
					O15440 homo sapien			
					P26660 hominis po			
					O19133 canis familiaris			
					P13892 schizosaccharomyces			
					P25001 pisaster oc			
					Ogere6 rattus norvegicus			
					O15440 homo sapien			
					P26660 hominis po			
					O19133 canis familiaris			
					P13892 schizosaccharomyces			
					P25001 pisaster oc			
					Ogere6 rattus norvegicus			
					O15440 homo sapien			
					P26660 hominis po			
					O19133 canis familiaris			
					P13892 schizosaccharomyces			
					P25001 pisaster oc			
					Ogere6 rattus norvegicus			
					O15440 homo sapien			
					P26660 hominis po			
					O19133 canis familiaris			
					P13892 schizosaccharomyces			
					P25001 pisaster oc			
					Ogere6 rattus norvegicus			
					O15440 homo sapien			
					P26660 hominis po			
					O19133 canis familiaris			
					P13892 schizosaccharomyces			
					P25001 pisaster oc			
					Ogere6 rattus norvegicus			
					O15440 homo sapien			
					P26660 hominis po			
	</							

CC OXXW motif and is present only in processive glycosyl transferases. It could be involved in the processivity function of the enzyme, possibly required for holding the growing glycan chain in the active site.

CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation in the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AE005579; AAG58675.1; ALT INIT. DR EMBL; AP002565; BAB37336.1; ALT INIT. DR InterPro; IPR001173; Glycos_transf_2. DR Pfam; PF00535; Glycos_transf_2. CC Cellulose biosynthesis; Transferase; glycosyltransferase; Transmembrane; Inner membrane; Complete proteome. KW DOMAIN 271 364 CATALYTIC SUBDOMAIN A. FT TRANSMEM 441 501 CATALYTIC SUBDOMAIN B. FT TRANSMEM 30 50 POTENTIAL. FT TRANSMEM 151 171 POTENTIAL. FT TRANSMEM 173 193 POTENTIAL. FT TRANSMEM 230 250 POTENTIAL. FT TRANSMEM 525 545 POTENTIAL. FT TRANSMEM 547 567 POTENTIAL. FT TRANSMEM 592 612 POTENTIAL. FT TRANSMEM 640 660 POTENTIAL. FT TRANSMEM 668 688 POTENTIAL. FT TRANSMEM 833 853 POTENTIAL. FT ACT_SITE 313 313 POTENTIAL. FT ACT_SITE 457 457 POTENTIAL. FT SITE 360 360 SUBSTRATE BINDING (POTENTIAL). FT SITE 362 362 SUBSTRATE BINDING (POTENTIAL). SQ SEQUENCE 872 AA; 99710 MW; F3F1A24A2B13FBA CRC64;

Query Match 4.8%; Score 273; DB 1; Length 872; Best Local Similarity 19.0%; Pred. No. 8.6e-12; Matches 153; Conservative 99; Mismatches 201; Indels 354; Gaps 32; Query 260 MWIMRLUVICLILHYRITNPV---PAFAFLWIVSVCIEIWAFALSWILDQFPKKWPWNR 315 201 MLTVLSLTVSCRYWTRSTLWDPPVSLVCGLLPLAETYAWIIVLIGYFQWVWPNR 260 316 ETYDRLALRYDRDREGEPSQALAVDIFVSVTDPKLEPPLVTANTVSLILAVDYPVDKVSCY 375 261 QP---VPLPKDMSLWPS---VDIFVPTN---EDLNVKNTIYASLGIDWPKDKLNW 309 376 VFDGGAMLSFEELAETSEFARKWWPEKKYSIEPRAPEWYFAAKIDYLKDQVQTSFYKD 435 310 ILDDGG-----315

QY 436 RRAMKREVEEKFIRINALVSKALKCPERCGWMODGTPMPEGNNNTGDDHPGMIQVLGLQNGGL 495 316 ---REERPRQF-----ANQNYGV 328

Db 496 DAEGNELPLRVVYSSREKRPQFQHHKACAGMNLAYRVSATLNGPFTILNUCDHYINNSKA 555 329 K-----VIART---HEHAKAGINNALYYA---KGBFVSDFCDAHPPTRSFL 370

QY 556 LRBAMCFELMDPNLGKQCVYQVFQRFQ---DGIDKN---DRYANRNTVFFDINLRGLCI 608 371 OMFGVWELKE---KOLAMMOTPHFFSPDPFERNLGRFRKTNEGTFYGYWQDDNMW 426 609 QGPVYVGTCVFNRTALGYEPPIKKVKRKPSLISKLCGGGSRARNKSKARKEKKDKKSRRH 668 427 DATFFCGSCAVIRR-----KP-----442

Db 669 TSTVTPVENDDEEFGVEGEGADDEKALLMSQMSLEKRFQGQSAVFASLTMENGGVPPSA 728 443 -----LDEL-----449

QY 729 TPENLKLAEHWIVISCGYEDKSDWGMIGWIGSVMDILEGPKHMARGWSISNCPLPA 788 450 -----AVE-----450

Db 789 FGKSAPINSLSRNLNOYLRWALGSVETLFSRHCPIWYGNR-LKLERFAVNTTYPIT 478 479 AGLAGATESLSGHIGORIRWAGMVOI-FRINPL---TAKGKFAQRCYVHAMEHLS 533 848 SIPLMCYCTLAVCLFTQNQFIPQINSIASIWLSLFLSFATGLEMNSGVIDEWMR 907 534 GIPRLFLFTAPLAFLHLAYTI---YAPLMLAFV--LPHMITHASLNTSKIOQKYR 585 908 NEOFWIGGVSIAHFLAVFOGLI-----VILAGLDTN---FTVSKASDEDGDFALYL 957 586 H-SFW-----SEIVETVLAWHTAPPVIALNPHKGKFNVPAK---GGLVEEY 630

Db 1018 MGRQNRPITPVVWSVLLASTFSLLW 1044 661 -----PPT-----EMLTWVSMWV 675

QY 958 FKWTTLIIPPTLTLIVNLGVAGYSAIANGSYOQSGPLGCKLFCAFFWVTHLYPFLKGL 1017 631 VDW-VISRPYTFVLVILNLGVAVGI-----WRYFYG-----660

Db 661 VDW-VISRPYTFVLVILNLGVAVGI-----WRYFYG-----660

QY 15-JUN-2002 (Rel. 41, Last sequence update) 15-JUN-2002 (Rel. 41, Last annotation update) Cellulose synthase catalytic subunit [UDP-forming] (EC 2.4.1.12). DE BCSCA OR Sty41B1. OS Salmonella typhi. OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella. OX NCBI_TaxID=601; RN [1]

RP ID BCSA_SALT1 STANDARD; PRT; 874 AA.

RC STRAIN=CT18;

RX MEDLINE=21534947; PubMed=11677608;

RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J., Churcher C., Mungall L., Bentley S.D., Holden M.T.G., Sebaiha M., Baker S., Basham D., Brooks D., Chillingworth T., Connerton F., Cronin A., Davis P., Davies R.M., Dowd D.L., White N., Farrar J., Fellwell T., Hamlin N., Hague A., Hien T.T., Holroyd S., Jaspels K., Krogh A., Larsen T.S., Leather S., Moulie S., O'Gaoar P., Parry C., Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.; "Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar *Typhi* CT18"; Nature 413:848-857(2001).

RL -1- FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes uridine 5'-diphosphate glucose to cellulose, which is produced as an extracellular component for mechanical and chemical protection at the onset of the stationary phase, when the cells exhibit multicellular behavior (rدار morphology). Co-expression of cellulose and thin aggregative fibrillae leads to a hydrophobic network with tightly packed cells embedded in a highly inert matrix (By similarity).

CC -1- CATALYTIC ACTIVITY: UDP-glucose + ((1,4-beta-D-glucosyl)(N)) = UDP + ((1,4-beta-D-glucosyl)(N)).

CC -1- COFACTOR: Magnesium (By similarity).

CC -1- ENZYME REGULATION: Activated by bis-(3'-5') cyclic diguanylic acid (c-di-GMP) (By similarity).

CC -1- PATHWAY: Bacterial cellulose biosynthesis.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane

CC -1- DOMAIN: There are two conserved domains in the globular part of the protein: the N-terminal domain (domain A) contains the conserved DXD motif and is possibly involved in catalysis and

CC substrate binding. The C-terminal domain (domain B) contains the
 CC QXRW motif and is present only in processive glycosyl
 CC transferases. It could be involved in the processivity function of
 CC the enzyme, possibly required for holding the growing glycan chain
 CC in the active site.

-1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR InterPro: IPR001173; Glycos_transf_2.

DR Pfam: PF0053; Glycos_transf_2.

KW Cellulose biosynthesis; Transferase; Glycosyltransferase;
 KW Transmembrane; Inner membrane; Complete proteome.

FT DOMAIN 271 364 CATALYTIC SUBDOMAIN A.

FT TRANSMEM 441 501 CATALYTIC SUBDOMAIN B.

FT TRANSMEM 30 50 POTENTIAL.

FT TRANSMEM 151 171 POTENTIAL.

FT TRANSMEM 173 193 POTENTIAL.

FT TRANSMEM 230 250 POTENTIAL.

FT TRANSMEM 252 255 POTENTIAL.

FT TRANSMEM 547 567 POTENTIAL.

FT TRANSMEM 592 612 POTENTIAL.

FT TRANSMEM 634 654 POTENTIAL.

FT TRANSMEM 668 688 POTENTIAL.

FT TRANSMEM 833 853 POTENTIAL.

FT ACT_SITE 313 313 POTENTIAL.

FT ACT_SITE 457 457 POTENTIAL.

FT SITE 360 360 SUBSTRATE BINDING (POTENTIAL).

FT SITE 362 362 SUBSTRATE BINDING (POTENTIAL).

SQ SEQUENCE 874 AA: MW: 33950D9CD0B51990 CRC64;

Query Match 4.8%; Score 273; DB 1; Length 874;
 Best Local Similarity 18.8%; Pred. No. 8.6e-12; Mismatches 201; Indels 370; Gaps 29;
 Matches 153; Conservative 91; Mismatches 201; Indels 370; Gaps 29;

OY 260 MVVIMRLVVLICLFHYRINTNPV---PNAFALWIVSVCIEIWFLSWILDQEPKWPVNR 315
 OY :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
 Db 201 MLTVISLTIVSCRYIWMRTYSTLNNDPVSLSVCGLTLFAETYWANIVLVLYGFOVWMLNIR 260
 OY 316 ETYIDLRLALRYDRIGEPPSOLAAVIFVSVDPKLKEPPLYTANTVLSILAVDYPDKVSCY 375
 OY :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
 Db 261 QP-----VPLPKEMSQWPPTVDIFVPTYN--EDLYNTVKNTIYASLGIDWMPDKLNUW 309
 OY 376 VFDDGAMLSFSELAETSFARKWNPCKYSIEPRAPEWYFAAKTDYLKDVKOTSFVWD 435
 Db 310 ILDDGGR----ESP----ROFAR---- 324
 OY 436 RRAMKREYEEFKIRINALVSKALKCPEEGWVMODGTPWGNNGDHPGMQLFGLQNGSL 495
 Db 325 -----YIAR----ATHEHAKGNINNALKAH----KGFRVATFDCDRVPTRSPL 370
 OY 496 DAEGNELPRLVYSREKRCFQHHKKAGAMNALYRVASVLTGPRILNLDCHDVINNSKA 555
 Db 330 -----YIAR----ATHEHAKGNINNALKAH----KGFRVATFDCDRVPTRSPL 370
 OY 556 LREAMCFMLDPNLGKQVQVQFOPRF---DGIDKN---DRYANRNTVFDINLRGLGDI 608
 Db 371 QMTNGWFLKE----KQLMQTPHFFSDPFERNLGRFRKTPTGELTYGLVQDGNDMW 426
 OY 609 QGPYVGTGCVFNRLHALGYVEPPRKVKHKRPSLSSKLCGGSRKNSKARKEKDRKSGRH 668
 Db 427 DATFFCGSCAVIRR----KP---- 442
 OY 669 TDSTVVPVNLDIIBEGVVEGAGFDDBKALLMSQMSLEKRGQSAVFVASTLMENGVPPSA 728
 Db 443 -----LDEI-----GGI---- 449

OY 729 TPENILIKEATHVIVSGCYEDKSDWGMEMIGWITGYSVTEDLTGKUHARGRSVYKMPKLPA 768
 OY :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
 Db 450 -----AVE----- 478
 Db 479 SAGLATESIUSAHQIQRIRWARGMVOI-FRLDNPL---FGKGKLQARCLQYLNAMPHFSLG 534
 OY 849 IPILMCT-----LLACLFNFIIOI-----SNIASTWFLSPLSTA 889
 OY :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
 Db 535 IPRILFLTAPLAELIHAYIYIAPALMIALFVPRINVHASNTRKIOGKYRHSMWEITE 594
 OY 950 GDFALSYLFKWTTLIPPTTLIVLWGVWAGVSYAISNGYQSGLPFLFLEFWIVH 1009
 OY :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
 Db 623 GGLVVEEKYVWD-VISRPYFLFLVLUMLIGVAVG-----WRYYYJ----- 661
 OY 1010 LYPFUKGLMRQNKRPIITVWWWSULASIFSLWV 1044
 OY :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
 Db 662 -----ENELTIVY-----SLWV 675

RESULT 3

BCSA_ECOLI ID BCSA_ECOLI STANDARD; PRT; 872 AA.
 AC P31653; P37654; P76712; P76713;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cellulose synthase catalytic subunit [UDP-forming] (EC 2.4.1.12).
 GN BCSA OR B3533.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=94310500; PubMed=8041620;
 RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
 RT Analysis of the Escherichia coli genome. V. DNA sequence of the
 RT region from 76.0 to 81.5 minutes".
 RL Nucleic Acids Res. 22:2576-2586(1994).
 RN [2]
 RP REVISIONS
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Iñiguez I., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Greco J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Ma B., Shao Y.;
 RT The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP CHARACTERIZATION.
 RC STRAIN=ECOR10, ECOR12, and TOBL;
 RX Published=11260463;
 RA Zogaj X., Niintz M., Rojahn M., Bokranz W., Roemling U.;
 CC The multicellular morphotypes of *Salmonella typhimurium* and
 CC *Escherichia coli* produce cellulose as the second component of the
 extracellular matrix.";
 RL Mol. Microbiol. 39:1452-1463(2001).
 CC -1- FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes
 CC uridine 5'-diphosphate glucose to cellulose, which is produced as
 CC an extracellular component for mechanical and chemical protection
 CC at the onset of the stationary phase, when the cells exhibit
 CC multicellular behavior (rدار morphology). Co-expression of
 CC cellulose and thin aggregative fibrillae leads to a hydrophobic
 CC network with tightly packed cells embedded in a highly inert

CC	-! CATALYTIC ACTIVITY: UDP-glucose + {(1,4-beta-D-glucosyl)}(N) = UDP matrix.	Db	310 ILDDGG-----	QY	436 RRAMKREYEFKIRINALVSKALKCPEEGWVMODGTPWPGNNNTGDPGMIVQFLGQNGL 495
CC	+ {(1,4-beta-D-glucosyl)}(N+).	QY	436 RRAMKREYEFKIRINALVSKALKCPEEGWVMODGTPWPGNNNTGDPGMIVQFLGQNGL 495	Db	316 --- REEFROF-----
CC	-! COFACTOR: Magnesium (BY similarity).	QY	496 DAEGELPRLVYVSREKRCFOHKKAGKAMNALVRSAVLINGPFLNUCDIYINNSKA 555	Db	329 K-----YIARTT---HEHAKGNNTNALKY---KGEPVSTFDCCAVPTRSFL 370
CC	-! ENZYME REGULATION: Activitated by bis-(3',-5') cyclic diguanylic acid (C-di-GMP).	QY	556 LREAMCFLMDPNLGKQCVQFFOFE--DGIDKN---DRYANRNTFFDILRGLGI 608	Db	371 QMTNGWFLKE---KQLAMMOTPHFFSDPFEEENLGRFRKTNEGTLFYGQDNDW 426
CC	-! DOMAIN: There are two conserved domains in the globular part of the protein, the N-terminal domain (domain A) contains the conserved DXD motif and is possibly involved in catalysis and substrate binding. The C-terminal domain (domain B) contains the OXXRW motif and is present only in processive glycosyl transferases. It could be involved in the processive function of the enzyme, possibly required for holding the growing glycan chain in the active site.	QY	609 QGPYVGTVGCVFNRLVYALGYEPPTRVKHKRKPSSLKLCGGSRKRNKSKAKKESDKKSGRH 668	Db	427 DATFCFGSCAVIRR-----KP-----
CC	-! MISCELLANEOUS: The genes bcsA, bcsB, bcsZ and bcsC are constitutively transcribed but cellulose synthesis occurs only when ada, a putative transmembrane protein regulated by agfd, is expressed. Cellulose production is abolished in E.coli K12.	QY	669 TDSTPVFVNLDLLERGVEDAFDEDDKALIMQSLEKFRGQSAVFWASTIMENGGVPPSA 728	Db	443 -----LDEI-----GGI----- 449
CC	-! SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2, frame shift in position 128.	QY	729 TPENILKEATHVWISCGYEDKSDWGMEMIGWIYGSVTEILDITGKMMHARGWSRSTCMKPLA 788	Db	450 -----AVE-----TWEDATSLRHRGRTSAM-RIFQ 478
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).	QY	789 FKGSAPINISDRMLNOVLRVALGSGVEILSRHCPIWYNGR-LKFLEFRAYVNTTIPIT 847	Db	479 AAGLATELSAHIGQIRRARGMVQI-FRLDNPL---TGKGLKFAQRLCYWNAMFHRLS 533
CC	-----	QY	848 SIPIMCYCTULLAVCFDTQFIPOINSNTASIWIPEFISIFANGILEMWPSGIDEWNR 907	Db	534 GIPRILFLTAPLAFLLHAYII---YAPALMIALFV---LPHMIHASLTISKIOGKR 585
CC	-----	QY	908 NEQFWWIGVSAHLAEFVGIL-----KVLAGIDTN---FWTSKASDEDGFELYL 957	Db	586 H-SFW-----SEIYEHVLANYIAPPFLVALINPHKGKENVAK-----GGIVEBEY 630
DR	Ecdone; EG1260; bcsA.	QY	958 FKWTTLLIPTTLVNLVNVGWWAGSYAINSGIQSGWGLPFLGKLFPAFWVIVHLYFLKL 1017	Db	631 VDW-VISRYFIVLVLNLVGVAGI-----WRYFYG----- 660
DR	InterPro: IPR01173; Glycos_transf_2.	QY	1018 MGROWRTPPTLVVWNSVLLASIFSLWV 1044	Db	661 -----PPT-----EMLTVVWSWW 675
KW	Cellulose biosynthesis; Transferase; Glycosyltransferase;	QY	586 H-SFW-----SEIYEHVLANYIAPPFLVALINPHKGKENVAK-----GGIVEBEY 630	Db	631 VDW-VISRYFIVLVLNLVGVAGI-----WRYFYG----- 660
FT	Transmembrane; Inner membrane	QY	958 FKWTTLLIPTTLVNLVNVGWWAGSYAINSGIQSGWGLPFLGKLFPAFWVIVHLYFLKL 1017	Db	661 -----PPT-----EMLTVVWSWW 675
FT	CATALYTIC SUBDOMAIN A.	QY	631 VDW-VISRYFIVLVLNLVGVAGI-----WRYFYG----- 660	Db	661 -----PPT-----EMLTVVWSWW 675
FT	DOMAIN 271 364	QY	1018 MGROWRTPPTLVVWNSVLLASIFSLWV 1044	Db	661 -----PPT-----EMLTVVWSWW 675
FT	POTENTIAL.	QY	586 H-SFW-----SEIYEHVLANYIAPPFLVALINPHKGKENVAK-----GGIVEBEY 630	Db	631 VDW-VISRYFIVLVLNLVGVAGI-----WRYFYG----- 660
FT	TRANSMEM 441 501	QY	958 FKWTTLLIPTTLVNLVNVGWWAGSYAINSGIQSGWGLPFLGKLFPAFWVIVHLYFLKL 1017	Db	661 -----PPT-----EMLTVVWSWW 675
FT	POTENTIAL.	QY	631 VDW-VISRYFIVLVLNLVGVAGI-----WRYFYG----- 660	Db	661 -----PPT-----EMLTVVWSWW 675
FT	TRANSMEM 151 171	QY	1018 MGROWRTPPTLVVWNSVLLASIFSLWV 1044	Db	661 -----PPT-----EMLTVVWSWW 675
FT	POTENTIAL.	QY	586 H-SFW-----SEIYEHVLANYIAPPFLVALINPHKGKENVAK-----GGIVEBEY 630	Db	631 VDW-VISRYFIVLVLNLVGVAGI-----WRYFYG----- 660
FT	TRANSMEM 173 193	QY	958 FKWTTLLIPTTLVNLVNVGWWAGSYAINSGIQSGWGLPFLGKLFPAFWVIVHLYFLKL 1017	Db	661 -----PPT-----EMLTVVWSWW 675
FT	POTENTIAL.	QY	631 VDW-VISRYFIVLVLNLVGVAGI-----WRYFYG----- 660	Db	661 -----PPT-----EMLTVVWSWW 675
FT	TRANSMEM 230 250	QY	1018 MGROWRTPPTLVVWNSVLLASIFSLWV 1044	Db	661 -----PPT-----EMLTVVWSWW 675
FT	POTENTIAL.	QY	586 H-SFW-----SEIYEHVLANYIAPPFLVALINPHKGKENVAK-----GGIVEBEY 630	Db	631 VDW-VISRYFIVLVLNLVGVAGI-----WRYFYG----- 660
FT	TRANSMEM 525 545	QY	958 FKWTTLLIPTTLVNLVNVGWWAGSYAINSGIQSGWGLPFLGKLFPAFWVIVHLYFLKL 1017	Db	661 -----PPT-----EMLTVVWSWW 675
FT	POTENTIAL.	QY	631 VDW-VISRYFIVLVLNLVGVAGI-----WRYFYG----- 660	Db	661 -----PPT-----EMLTVVWSWW 675
FT	TRANSMEM 547 567	QY	1018 MGROWRTPPTLVVWNSVLLASIFSLWV 1044	Db	661 -----PPT-----EMLTVVWSWW 675
FT	POTENTIAL.	QY	586 H-SFW-----SEIYEHVLANYIAPPFLVALINPHKGKENVAK-----GGIVEBEY 630	Db	631 VDW-VISRYFIVLVLNLVGVAGI-----WRYFYG----- 660
FT	TRANSMEM 592 612	QY	958 FKWTTLLIPTTLVNLVNVGWWAGSYAINSGIQSGWGLPFLGKLFPAFWVIVHLYFLKL 1017	Db	661 -----PPT-----EMLTVVWSWW 675
FT	POTENTIAL.	QY	631 VDW-VISRYFIVLVLNLVGVAGI-----WRYFYG----- 660	Db	661 -----PPT-----EMLTVVWSWW 675
FT	TRANSMEM 640 660	QY	1018 MGROWRTPPTLVVWNSVLLASIFSLWV 1044	Db	661 -----PPT-----EMLTVVWSWW 675
FT	POTENTIAL.	QY	586 H-SFW-----SEIYEHVLANYIAPPFLVALINPHKGKENVAK-----GGIVEBEY 630	Db	631 VDW-VISRYFIVLVLNLVGVAGI-----WRYFYG----- 660
FT	TRANSMEM 668 688	QY	958 FKWTTLLIPTTLVNLVNVGWWAGSYAINSGIQSGWGLPFLGKLFPAFWVIVHLYFLKL 1017	Db	661 -----PPT-----EMLTVVWSWW 675
FT	POTENTIAL.	QY	631 VDW-VISRYFIVLVLNLVGVAGI-----WRYFYG----- 660	Db	661 -----PPT-----EMLTVVWSWW 675
FT	ACT_SITE 833 853	QY	1018 MGROWRTPPTLVVWNSVLLASIFSLWV 1044	Db	661 -----PPT-----EMLTVVWSWW 675
FT	POTENTIAL.	QY	586 H-SFW-----SEIYEHVLANYIAPPFLVALINPHKGKENVAK-----GGIVEBEY 630	Db	631 VDW-VISRYFIVLVLNLVGVAGI-----WRYFYG----- 660
FT	ACT_SITE 313 313	QY	958 FKWTTLLIPTTLVNLVNVGWWAGSYAINSGIQSGWGLPFLGKLFPAFWVIVHLYFLKL 1017	Db	661 -----PPT-----EMLTVVWSWW 675
FT	POTENTIAL.	QY	631 VDW-VISRYFIVLVLNLVGVAGI-----WRYFYG----- 660	Db	661 -----PPT-----EMLTVVWSWW 675
FT	SEQUENCE 360 457	QY	1018 MGROWRTPPTLVVWNSVLLASIFSLWV 1044	Db	661 -----PPT-----EMLTVVWSWW 675
FT	SITE 360 362	QY	586 H-SFW-----SEIYEHVLANYIAPPFLVALINPHKGKENVAK-----GGIVEBEY 630	Db	631 VDW-VISRYFIVLVLNLVGVAGI-----WRYFYG----- 660
FT	SUBSTRATE BINDING (POTENTIAL).	QY	958 FKWTTLLIPTTLVNLVNVGWWAGSYAINSGIQSGWGLPFLGKLFPAFWVIVHLYFLKL 1017	Db	661 -----PPT-----EMLTVVWSWW 675
SQ	SEQUENCE 362 362	QY	631 VDW-VISRYFIVLVLNLVGVAGI-----WRYFYG----- 660	Db	661 -----PPT-----EMLTVVWSWW 675
Query Match	4.8%; Score 272; DB 1; Length 872; Best Local Similarity 19.0%; Pred. No. 1e-11; Indels 354; Gaps 32; Matches 153; Conservative 99; Mismatches 201; Indels 354; Gaps 32;	RESULT 4	BCSA_SALTY	BCSA_SALTY STANDARD; PRT; 874 AA.	
OY	260 MVIMRLVILVCLFLHYRTNPV---PNAFLWVSVCEIWFALWILDQFRKWPVNR 315	ID	Q93IN2;	AC	Q93IN2;
Db	201 MLIYLSLTWCRYIWWVYIYSTLNDDPVSILWVCGILLFRFETAYAVIVLVLGYFQVWPLNR 260	DT	15-JUN-2002 (Rel. 41, Created)	DT	15-JUN-2002 (Rel. 41, Last sequence update)
OY	316 ETYLDRLALRYDREGEPSCOLAVIDFVSTWDPLKEPLPTANTVLSILADVYPVDKVSYY 375	DT	15-JUN-2002 (Rel. 41, Last annotation update)	DE	Cellulose synthase catalytic subunit [UDP-forming] (EC 2.4.1.12).
Db	261 QP---VPLPKDMSLWPS---VDFVPPYN---EDLNVKNTIYASLGIDWPKDLNWI 309	GN	BCAA OR STM3119.	OS	(Salmonella typhimurium).
OY	376 VFDGAAMSFESLAETSEFAKRWVPCFKYSEPRAEWVFAAKIDYLKDQVTSFWKD 435	OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella.	NCI	NCI_TaxID=602;
QX	STRAIN=TAXID=602;	RN	SEQUENCE FROM N.A.	RC	SEQUENCE FROM N.A.
QX	STRAIN=ETCC 14028;	RN	STRAIN=ETCC 14028;	RC	STRAIN=ETCC 14028;
QX	MEDLINE=21160181; PubMed=11260463;	RN	MEDLINE=21160181; PubMed=11260463;	RC	MEDLINE=21160181; PubMed=11260463;
ZY	Zogaj X., Rohde M., Boekranz W., Roemling U.; "The multicellular morphotypes of <i>Salmonella typhimurium</i> and <i>Escherichia coli</i> produce cellulose as the second component of the extracellular matrix." <i>J. Microbiol.</i> 39:1452-1463 (2001).	RN	[2]	RN	SEQUENCE FROM N.A.
QX	SIRAIN=L12;	RC	MEDLINE=21927388; PubMed=11929533;	RC	MEDLINE=21927388; PubMed=11929533;

RA Solano C., Garcia B., Valle J., Berasain C., Ghigo J.-M., Gamazo C.,
 RA Lasa I.;
 RT "Genetic analysis of *Salmonella enteritidis* biofilm formation:
 RT critical role for cellulose."
 RL Mol. Microbiol. 43:793-808(2002).
 RN [3]
 SEQUENCE FROM N. A.
 RC STRAIN=LT2 / SS3C1412 / ATCC 700720;
 RX MEDLINE=200334948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latrelle P.,
 Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
 LT2."
 RL Nature 413:852-856(2001).
 CC -I- FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes
 uridine 5'-diphosphate glucose to cellulose, which is produced as
 an extracellular component for mechanical and chemical protection
 at the onset of the stationary phase, when the cells exhibit
 multicellular behavior (rod morphotype). Co-expression of
 cellulose and thin aggregative fimbriae leads to a hydrophobic
 network with tightly packed cells embedded in a highly inert
 matrix.
 CC -I- CATALYTIC ACTIVITY: UDP-glucose + {(1,4-beta-D-glucosyl)(N) - UDP
 + [(1,4-beta-D-glucosyl)](N+1)}
 CC -I- COFACTOR: Magnesium (By similarity).
 CC -I- ENZYME REGULATION: Activated by bis-(3',5') cyclic diguanylic acid
 (G-di-GMP) (By similarity).
 CC -I- PATHWAY: Bacterial cellulose biosynthesis.
 CC -I- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 (Potential).
 CC -I- DOMAIN: There are two conserved domains in the globular part of
 the protein: the N-terminal domain (domain A) contains the
 conserved DXD motif and is possibly involved in catalysis and
 substrate binding. The C-terminal domain (domain B) contains the
 QXKR motif and is present only in processive glycosyl
 transferases. It could be involved in the processivity function of
 the enzyme, possibly required for holding the growing glycan chain
 in the active site.
 CC -I- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC

CC

FT	ACT_SITE	313	313	POTENTIAL.
FT	ACT_SITE	457	457	POTENTIAL.
FT	SITE	360	360	SUBSTRATE BINDING (POTENTIAL).
FT	SEQUENCE	362	362	SUBSTRATE BINDING (POTENTIAL).
SQ	SEQUENCE	874 AA:	10043 MW:	AC9421B58606310A CRC64;
QY	Query Match	4.8%	Score 271;	DB 1; Length 874;
	Best Local Similarity	18.8%	Pred. No. 1.	2e-11;
	Matches	153;	Conservative	90;
			Mismatches	202;
			Indels	370;
			Gaps	29;
Db	260	MVIMMLRLVILCLFLHYRITNPV---PNAFAFLWLVSVICETMFAWSNLDOOPPKWFVN 315		
QY	201	MLIVLSLTIVSCRIVIWWRYITSLNWDDSVLWGLLFAETYAWIWLVLGFQVWPLNR 260		
QY	316	ETVLDRILALRVYREGERSQALAVDIFVSTVDPEKEPLPVANTVLSLAVDVWPDVKSCY 375		
Db	261	QP-----VPLPKEMQSOWPTVDFVPTYN--EDLNVKVNNIYASLGIDPMKDKLNW 309		
QY	376	VFDGAAMLSFESLAETSEFARKWNVFCKKYSIEPRAPEWFAAKIDYLKDQVQTFSV 435		
Db	310	ILDQGR----BSF--RHFA-----		
QY	436	RRAMKREYEFKIRRNALVSKALKCPSEGWWMODGTPWPGNNNGDHGMQIVFLGONGGL 495		
Db	325	-----HVGVH-----		
QY	496	DAEGNELPRLVVYVSEKRPGEFHKKAGAMNALVYRUSAVLTLNGPFILNLDCHDHYINNSKA 555		
Db	330	-----YIARTT---HEHAKAGNNINAKHA---KGEEFVAIFDCDHRVPTRSF 370		
QY	556	LREAMCFLMDPNUKGQVYQVFPQRF--DGIDKN---DRYANRNFEDINLRGLDGI 608		
Db	371	OMTMGWELKE---KOLAMMOPHFFSPDPFERNLNGFRKFPEGSTLFYGLVODGMW 426		
QY	609	QGPVYVGTCVENVRTALYGYEPPKVKVHHKPSLLSKLGGSRKNSKAKKESDKKSGRH 668		
Db	427	DATFFCGSCAVIRR-----KP-----		
QY	669	TDSTVPUVNLDDIEEGVEGAGFDDDEKALLMSOMSLERKGQSAVFVASTLMENGVP 728		
Db	443	-----LDEI-----		
Db	449	-----GIG-----		
QY	729	TPENLKEIAHVIVSGYEDKSDWMEIGWIYGSVTEIDLIGKMMHARGWRSTYCMPLKPA 788		
Db	450	-----AVE-----		
QY	789	FKGSSAPINLSDRUNQVLAWSVLEIFSRHCPIWYGRKLFERRAYVNNTIYTIS 848		
Db	479	AAGLATESLSAHIGQTRWARGAVQI FRDNP-----FGKGKIKLQAQLCYLYNAMFHRLSG 534		
QY	849	IPLMLMYCT-----LLAVCLFTNOFIPQI-----SNTIASIWELSLSIFA 889		
Db	535	IPLRLIFTAPLAFLFLHAYIYAPPALMTALFYVPHMVHASLTNSKICKYRHSFWSEYE 594		
QY	890	TGLEMMRMSGVGEDEWRNEQFWVIGGVSAALEFOGQILKULAGITNTFTK RASDED 949		
Db	595	TVLA-----WYIAPBT-----LVALINPHKGKENTK----- 622		
QY	950	GDFDAELYFLPKWTTLLIPPTLLVNLVNVGAGSYAINTSGYQSWGPFLGALFFAWVIVH 1009		
Db	623	GGVLEEKVWD-VISRPIVFLVLLNLGLGVAVG-----WRYVIG----- 661		
QY	1010	LYPFELKGLMGROMRNPPTIVVWVSLASIFSLWV 1044		
Db	662	-----ENELTVIV-----SLVWV 675		
RESULT	5			
BCSA_XANAC	ID	BCSA_XANAC	STANDARD;	PRT;
AC	P58932;			729 AA.
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			

DT		15-JUN-2002 (Rel. 41, Last annotation update)		Cellulose synthase catalytic subunit [UDP-forming] (EC 2.4.1.12).		SUBSTRATE BINDING (POTENTIAL).	
DN		BCSA OR XAC518.		Xanthomonas axonopodis (pv. citri)		BCS464;	
OS		bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;		Xanthomonas		NCBI_TaxID:92829;	
RN	RP	SEQUENCE FROM N.A.	SBQQUENCE FROM N.A.	PREDICTION	PREDICTION	SCORE	LENGTH
RC	STRAIN=306 / ATCC 13902 / XV 101;	252.5	DB 1;	1	Length 729;		
RX	MEDLINE=2202145; Published=202417;						
RA	Quaggio R.B., Monteiro-Vitorello C.B., Van Sluis M.A., Almeida N.F., Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A., Camarotte G., Caamaño F., Cardozo J., Chambargo F., Cipripa L.P., Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H., Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T., Formighieri E.M.C., Greggio C.C., Gruber A., Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F., Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M., Martins E.C., Medeiros J.J., Menck C.M., Miyake C.Y., Moon D.H., Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R., Pereira H.A., Rossi A., Senna J.A.D., Silva C., de Souza R.F., Spinola L.A.F., Takita M.A., Tanura R.E., Teixeira E.C., Tezza R.I.D., Trindade dos Santos M., Truffi D., Tsai S.M., White F.F., RT	4.4%	Score 252.5; DB 1; Length 729;				
RL	"Comparison of the genomes of two Xanthomonas pathogens with differing host specificities."						
RT	Nature 417:459-463(2002)						
CC	-1- FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes uridine 5'-diphosphate glucose to cellulose, which is produced as an extracellular component for mechanical and chemical protection (By similarity).						
CC	-1- CATALYTIC ACTIVITY: UDP-glucose + [(1,4-beta-D-glucosyl)(N) - UDP + (1,4-beta-D-glucosyl)](N) = UDP						
CC	-1- ENZYME REGULATION: Activated by bis-(3'-5') cyclic diguanylic acid (cGMP) (By similarity).						
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane (Potential).						
CC	-1- DOMAIN: There are two conserved domains in the globular part of the protein: the N-terminal domain (domain A) contains the conserved DXD motif and is possibly involved in catalysis and substrate binding. The C-terminal domain (domain B) contains the OXXRW motif and is present only in processive glycosyl transferases. It could be involved in the processivity function of the enzyme. Possibly required for holding the growing glycan chain in the active site.						
CC	-1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.						
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).						
CC	DR EMBL: AE012000; AACM8361.1; KW Cellulose biosynthesis; Transferase; Glycosyltransferase; Transmembrane; Inner membrane.						
FT	DOMAIN 151 244 CATALYTIC SUBDOMAIN A.	4.4%	POTENTIAL.				
FT	DOMAIN 321 381 CATALYTIC SUBDOMAIN B.	4.4%	POTENTIAL.				
FT	TRANSMEM 30 50 POTENTIAL.	4.47	POTENTIAL.				
FT	TRANSMEM 110 130 POTENTIAL.	5.40	POTENTIAL.				
FT	TRANSMEM 171 191 POTENTIAL.	5.49	POTENTIAL.				
FT	TRANSMEM 405 425 POTENTIAL.	6.10	POTENTIAL.				
FT	TRANSMEM 427 447 POTENTIAL.	6.30	POTENTIAL.				
FT	TRANSMEM 520 540 POTENTIAL.	6.30	POTENTIAL.				
FT	TRANSMEM 549 569 POTENTIAL.	6.30	POTENTIAL.				
FT	TRANSMEM 610 630 POTENTIAL.	6.30	POTENTIAL.				
ACT SITE	193 193 POTENTIAL.	6.30	POTENTIAL.				
OX	NCBI_TaxID:92829;						
RN	[1] SSEQUENCE FROM N.A.						
RP	SBQQUENCE FROM N.A.						
QY	260 MIVMLRVLVCLFLHYRITNP--VPNA--FALMLVVICETWFALSILQDFEKFWMFVN 315						
Do	81 MMGGMMSLAVSCLRVTWWRMTOPTMGVSADVIFSTVDLKEPLVTTAVLISLAVLDPV 140						
QY	316 ETYFLDRALRVREGEPSQLADIFVTFDLSKLEPLVTTAVLISLAVLDPV 375						
Db	141 KP---VPLPAQDRLWPS---VDVFIFTYN--EPISWVRTVLAASVDPAGKITH 189						
QY	376 VFDDGAAMLSFESLAESEFAKRWWPCKKYSIEPRAPEWYFAAKIDYLKDQVTSFVKD 435						
Db	190 LLDG----- 194						
Db	208 -----INVVTRTNA---HAKAGNNIAALK---KCSDGYVAIFDCDH-IPTRSF 249						
QY	556 LREAM-CFLMDPPNIGKQVCYVQFPQRF-----DGIDKNDYANRNTVFEFLINLGDLQ 607						
Db	250 LOVAMGWFHLDTKL---ALWOMPHYFSPDPDFERNLTDHGKVPNEGELFYGLQDNDQ 305						
QY	608 IQGPVYVGTGVCFNRTALYGFPEPIKVKHHKKPSLLSKLCGGSSRKNSKAKKEDKKSGR 667						
Db	306 WMAFFGCCSCAVIKTA----- 322						
QY	496 DAEGNEGLPRLVVYSREKRPGFQDHKKAGAMNALVRVSAVLTGPFILNDQCDHVNNSKA 555						
Db	208 -----INVVTRTNA---HAKAGNNIAALK---KCSDGYVAIFDCDH-IPTRSF 249						
QY	608 IQGPVYVGTGVCFNRTALYGFPEPIKVKHHKKPSLLSKLCGGSSRKNSKAKKEDKKSGR 667						
Db	306 WMAFFGCCSCAVIKTA----- 322						
QY	668 HTDSTVPVFNLDIEGVEGAGFDDDEKALLMSQMSLERFGOSAVFVASTKENGGPPS 727						
Db	323 ----- 329						
QY	728 ATPENLILKEAIHVISCGYEDKSDWGMELGWIYGSVTEILDGFKHMARGWRSIYCHPKLP 787						
Db	330 -----AVER----- 357						
QY	788 AFKGSAPINLSDRINQVLRWALGSVET----LFSRHCPIWGYGNRLKTEFAYVNT 841						
QY	901 GDDEMRWNRNEQFWIGVSAHLAVFOGILKVIAGID-NFTWTSKA-----SDEDDEAE 954						
Db	358 QAAAGLATESLIGHVAQIRWANGMAQIARIDNPLGR-----GLKLSQRLYLNAA 407						
QY	842 TYPITSPILLMCTILAVLCTNQFTIPOISNIASIWFLSFLSFATG-----LEMRSVG 900						
Db	408 MHFFYGVPRILYLTAPLAYLFGAHVIQASLMILAYALPHIQLQNLTLNEVOSRRHL 467						
QY	901 GDDEMRWNRNEQFWIGVSAHLAVFOGILKVIAGID-NFTWTSKA-----SDEDDEAE 954						
Db	468 LNNEVYVETLWAVT-----FRPLVALNPKLGFNTVPKGGLVARSYDAOIAK 517						
QY	955 LVIKFWKTTLLPPTTLIVLNLYGWVAGY--SWAINSGYQS-WGPLEGKLFPAFWIVH 1009						
Db	518 PYLF-----LLLNVVMGAVLRLIVGGSGEQTIW-----ENLANTLYN 559						
RESULT 6							
AC52_ACExY							
ID	AC52_ACExY STANDARD;						
AC	059167;						
DT	15-JUN-2002 (Rel. 41, Created)						
DT	15-JUN-2002 (Rel. 41, Last sequence update)						
DT	15-JUN-2002 (Rel. 41, Last annotation update)						
FT	SEQUENCE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	SITE 240 242 242						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	SEQUENCE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						
FT	SITE 240 240						
FT	SITE 729 AA; 80915 MW; B9C0BB9595E795B1						
FT	ACT_SITE 337 337						

CC -I PATHWAY: Bacterial cellulose biosynthesis.
 CC -I SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (POTENTIAL).
 -I DOMAIN: There are two conserved domains in the globular part of
 the catalytic subunit: the N-terminal domain (domain A) contains
 the conserved DxD motif and is possibly involved in catalysis and
 substrate binding. The C-terminal domain (domain B) contains the
 QXKRW motif and is present only in processive glycosyl
 transferases. It could be involved in the processivity function of
 the enzyme, possibly required for holding the growing glycan chain
 in the active site.
 -I SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
 GLYCOSYLTRANSFERASE FAMILY 2.
 -I SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ACB/BCCS
 FAMILY.
 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 EMBL: AB015804; BAA77601; -;
 InterPro: IPR003919; CellSynth_A.
 InterPro: IPR00920; CellSynth_B.
 InterPro: IPR00173; Glycos_transf_2.
 Pfam: PF00355; Glycos_transf_2; 1.
 PRINTS: PRO1439; CELSNTHASEA.
 PRINTS: PRO1440; CELSNTHASEB.
 Cellulose biosynthesis; Transferase; Glycosyltransferase;
 Transmembrane; Inner membrane.
 DOMAIN 1 731 CATALYTIC.
 DOMAIN 2 1518 CYCLIC DI-GMP BINDING (BY SIMILARITY).
 DOMAIN 237 CATALYTIC SUBDOMAIN A.
 DOMAIN 314 374 CATALYTIC SUBDOMAIN B.
 PRT TRANSEM 24 44 POTENTIAL.
 PRT TRANSEM 71 91 POTENTIAL.
 PRT TRANSEM 105 125 POTENTIAL.
 PRT TRANSEM 144 237 POTENTIAL.
 PRT TRANSEM 314 374 POTENTIAL.
 PRT TRANSEM 465 485 POTENTIAL.
 PRT TRANSEM 514 534 POTENTIAL.
 PRT TRANSEM 543 563 POTENTIAL.
 PRT TRANSEM 1481 1501 POTENTIAL.
 PRT ACT_SITE 186 186 POTENTIAL.
 PRT ACT_SITE 330 330 POTENTIAL.
 SITE 233 233 POTENTIAL FOR SUBSTRATE BINDING
 (POTENTIAL).
 SITE 235 235 IMPORTANT FOR SUBSTRATE BINDING
 (POTENTIAL).
 SEQUENCE 1518 AA; 166464 MW; 7D634503183DAB6 CRC64;
 SQ RESULT 9
 BCA2_ACExY
 ID BCA2_ACExY STANDARD; PRT; 756 AA.
 AC 082859;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DB Cellulose synthase catalytic subunit [UDP-forming] (EC 2.4.1.12).
 GN BCSA.
 OS Acetobacter xylinus.
 OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
 OC Gluconacetobacter.
 OX NCBI_TaxID:28448;
 RN [1]
 RP SEQUENCE FROM N A.
 RC STRAIN=BPR 2001;
 RX MEDLINE=98396257; PubMed=9630529;
 RA Nakai T., Moriya A., Tonouchi N., Tsuchida T., Yoshinaga F.,
 RA Horimouchi S., Sone Y., Mori H., Sakai F., Hayashi T.;
 RT Control of expression by the cellulose synthase (BCSA) promoter
 RT region from Acetobacter xylinum BPR 2001.;
 RL Gene 213:95-100(1998).
 CC -I FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes
 uridine 5'-diphosphate glucose to cellulose. The thick cellulosic
 mats generated by this enzyme probably provide a specialized
 protective environment to the bacterium (By similarity).
 -I CATALYTIC ACTIVITY: UDP-glucose + [(1,4)-beta-D-glucosyl](N) = UDP
 + [(1,4)-beta-D-glucosyl](N+1).
 CC -I COFACTOR: Magnesium (By similarity).
 CC -I ENZYME REGULATION: Activated by bis-(3'-5') cyclic diguanosylic acid
 (C-di-GMP).
 CC -I PATHWAY: Bacterial cellulose biosynthesis.
 CC -I SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 (Potential).
 CC -I DOMAIN: There are two conserved domains in the globular part of
 the entry.
 Db 207 -PAENN-----HAKAGNLNLHALV---TDSPPAVIFDCDH-VPTRG 2411
 Qy 555 ALREAMCFLM-DPNLGKQVCYVQPQRDGKKNDRYANRNTVFFDINLRGLGIGQVY 653
 Db 242 FLRTTICWMMADPNL---ALIQTPHRYAPP-----FORNLAG--GMHVP-- 2828
 Qy 614 VGTGCVNVRNTALYGYEPEPIKVHKKPSLLSKCGGSRKNSKAKKESDKKSQRHTDSTV 6733
 Db 283 -PEGNMF----YGL----- 2916
 Qy 674 PVENLDDIEEGEGEAGETDEDEKAALLMSOMSLERKGFSVAVFVASTLMEENGVPPSATPNL 733
 Db 292 -----VQDGND----- 3111
 Qy 734 LKEAIHWISCGVEDKSDMGME-GWIVSVTEBILTGFMHARGWSRSTYCMPLKPAFKSA 7933
 Db 312 RHEAVMIGG-----GFATEHVTDAAHTALKMORRGWTAYLREPLAA-GLA 3566
 Qy 794 PNLSDRNQNRVWAALGSVEILFSRHSRHPWIYQNGRKFLEFRAYVNNTIYPITSIPIM 853
 Db 357 TERLILHIGQRWRWARGMIOQIRLDNPMLG---LRWEQRLCYLSAMSHEFAIRLT 4122
 Qy 854 YCTLLAVCLFTPHOFITPOISNTASITWL-SLUSISFANGILRNBRWSGVGIDEWNRNCFW 9122
 Db 413 FTVSPLAFLFLGQNLITIASPLATSVALPHIRHSVTSRIGRW-----RVSFW 4622
 Qy 913 -VIGGVS AHLFVAFQGILKVLAGIDTNFTVTSKASDEGDFFAEYLKWTILLPITT 9711
 Db 463 SEIYENSLAFLFVRYITVTLQPHKGKPNVT---DRKGLLARY-FWDA-VYPNVILA 5166
 Qy 972 IVNLVGVWAGVSYAINSQYQSMGFLGKLFFFAWFWVIVLHYPFLKL-MGRQR 1023
 Db 517 GWCAALLRGV-FGIWQFHDLRQLAQSPFLNLTWWVISLIVIASIANGRET 568

CC	the catalytic subunit; the N-terminal domain (domain A) contains the conserved DxD motif and is possibly involved in catalysis and substrate binding. The C-terminal domain (domain B) contains the QXXW motif and is present only in processive glycosyltransferases. It could be involved in the processivity function of the enzyme, possibly required for holding the growing glycan chain in the active site.
CC	-!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).
CC	-----
DR	EMBL: AB010645; BRA31463.1; -
DR	InterPro: IPR003919; CellSynth_A;
DR	Pfam: PF00535; Glycos_transf_2.
DR	PRINTS: PR01439; CELLSWHASEA.
KW	Cellulose biosynthesis; Transferase; Glycosyltransferase; transmembrane; Inner membrane.
FT	DOMAIN 147 242 CATALYTIC SUBDOMAIN A.
FT	DOMAIN 319 379 CATALYTIC SUBDOMAIN B.
FT	TRANSMEM 27 47 POTENTIAL..
FT	TRANSMEM 49 69 POTENTIAL..
FT	TRANSMEM 106 126 POTENTIAL..
FT	TRANSMEM 167 187 POTENTIAL..
FT	TRANSMEM 409 429 POTENTIAL..
FT	TRANSMEM 432 452 POTENTIAL..
FT	TRANSMEM 470 490 POTENTIAL..
FT	TRANSMEM 517 537 POTENTIAL..
FT	TRANSMEM 551 571 POTENTIAL..
FT	ACT_SITE 189 189 POTENTIAL..
FT	ACT_SITE 335 335 SUBSTRATE BINDING (POTENTIAL)..
FT	SITE 238 238 SUBSTRATE BINDING (POTENTIAL)..
FT	SITE 240 240 SUBSTRATE BINDING (POTENTIAL)..
SO	SEQUENCE 756 AA: 84561 MW: 6954F39A25E73B0A CRC64;
Query Match Similarity 4.2%; Score 236.5; DB 1; Length 756; Matches 166; Conservative 98; Mismatches 263; Indels 331; Gaps 34; Pred. No. 3.1e-09;	
QY	204 EKNIGPVSTQASBERGGVVDASD---ILADBDALLNEAR-----OPL 244
Db	11 ESRIGRISKILSLURGASYIVGALGLCALLAATVTLNNEOLIVAACVAVFFVVGKGRK 70
QY	245 SRKVSIPISSRINPYRMVITMLRLVTCFLHYRITNPVNPAFLWL----VSVICEIW 298
Db	71 SRRYQI-----FLEVLSALWSLRVLTWRLTETLD--FNTWIQGIGLVILMAELA 119
QY	299 ALSVILDDEPPKKWPNRTEYLDLALRDRREGPSQLAAVDIVSTVPLKEPLVANT 358
Db	120 LYMLFLSYPTQIOPHLRAB---LPPLPNDMPT---WDTEIPTYD---RDLSTVRL 168
QY	359 VLSILAVDVKYVDDGAMLSFESLAEATSEFARKWVPECKYSIEPRAPEWFA 418
Db	169 VLGALGIWMPDPKVNVYIDG----- 190
QY	419 AKIDYKLDKVOTSFVKDRRAMKREYEFKRIRINALVSKALKCPEEGWMDGPWPWGNNT 478
Db	191 -----SAHAKAGNINHAIK---MSG 231
QY	479 GDHPGMTOFELGGGLDAGEGNEPLRPLVYVSRERKRGPGQFHKKAGAMMALVRVSAVLTG 538
Db	203 ---GALYT-----GRVDF----- 202
QY	539 PFIINLDCDHYINNSKALRERAMCFMLDPNLKGKOVYQFPQRDGIDKNDRYANRNTVF 598
Db	232 DYLILDDCH-ITRAFLQIAWMW---ADRKIALMOTPHARYSPDP----- 275
QY	599 DINURGLDQIOPVYVGCGCVENTALGYEPPIKKHKKPSULSKLGGSRSRKNRKK 658
Db	276 -----FORNLAVGYRT----- 287
QY	659 ESDKKKSGRHTDSIVPVNLDDTEEGVEGAGEDDEKALLMSQMSLERKFGQSAVFVASTL 718
Db	288 -----PEGNL--FYGVTDGND-----FWDAFFCCSCA 314
QY	719 MENGGVPPSATPENLKEAHIVTSCGYEDKSDWGMELCIWIGSVTEDILTGFKMRHWR 778
Db	315 I-----LRREAESIG-----GFAEVTEAHTALRMQRGW 348
QY	779 SIYCMKPULKAFKQASAPINLSDRNQNOVLUWALSVETFSRHCIIWYGRN 837
Db	349 TAYL--RIFVASGLATERLTHIQRMWARGMQLI-FRVDNPML---GRGKLGORLC 401
QY	838 YVNNTTIVPITSIPLMYCTLLAVCLFTNQFLIPQINSIASIPLSLELSIFNGILEMRW 897
Db	402 YLSMTSFEFAIRVIFLASPLAFLAQNITAAPLAVALAVALPHMFHSIAI----- 454
QY	898 SGVQIDEWARNEQTW-VIGGSVASHLFAYFQGILKVLAGIDTMTVTWSKA---SDEDGFA 953
Db	455 -AAKVNKGR-YSEWSEYETTMALFLYRVTIVTLFLPSKGKENVTEKGGVLEEEIDLG 512
QY	954 ELY-----LFKWHITLIPPTPLI-VNENGWVAGVSTAISNGYQSMGPLEGKFLEAFWV 1006
Db	513 ATYPNIFATIMMGGLLGLFELIVRFNOLDVIARNAYLNCA-----WA 557
QY	1007 IVHYPFLKGL-MGRQR 1023
Db	558 LISLILFAIAVGRETK 575
RESULT 10	
ID	BCA1_AC0XY
BCA1_AC0XY	STANDARD;
PRT	754 AA.
AC	P19449;
DT	01-FEB-1991 (Rel. 17, Created)
DT	01-FEB-1991 (Rel. 17, Last sequence update)
DT	15-JUN-2002 (Rel. 41, Last annotation update)
DE	Cellulose synthase catalytic subunit [UDP-forming] (EC 2.4.1.12).
GN	BESA.
OS	Acetobacter xylinum.
OC	Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae; Gluconacetobacter.
OX	NCBI_TAXID=28448;
[1]	SEQUENCE FROM N.A., AND SEQUENCE OF 25-42 AND 196-206.
RC	STRAIN=1306-3;
RX	MEDLINE-91045951; Pubmed=2146681;
RA	Wong H.C., Fear A.L., Calhoun R.D., Eichinger G.H., Mayer R., Amitkam D., Ben-Ziman M., Gefland D.H., Meade J.H., Emerick A.W., Bruner R., Ben-Bassat A., Tal R.; "Genetic organization of the cellulose synthase operon in Acetobacter xylinum"; Proc. Natl. Acad. Sci. U.S.A. 87:8130-8134(1990).
RN	[2] ENZYME REGULATION.
RP	STRAIN=1306-3;
RX	PubMed=1297407;
RA	Chang A.L., Tuckerman J.R., Gonzalez G., Mayer R., Weinhouse H., Volman G., Amitkam D., Ben-Ziman M., Gilles-Gonzalez M.-A.; "Phosphodiesterase A1, a regulator of cellulose synthesis in Acetobacter xylinum, is a heme-based sensor.";
RT	"The thick cellulosic mats generated by this enzyme probably provides a specialized protective environment to the bacterium."
RL	Biochemistry 40:3420-3426(2001).
CC	-!- FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes uridine 5'-diphosphate glucose to cellulose in a processive way.
CC	The thick cellulosic mats generated by this enzyme probably provides a specialized protective environment to the bacterium.
CC	-!- CATALYTIC ACTIVITY: UDP-glucose + {(1,4)-beta-D-glucosyl}(N) - UDP + {(1,4)-beta-D-glucosyl}(N).
CC	-!- COFACTOR: Magnesium.
CC	-!- ENZYME REGULATION: Activated by bis-(3'-5') cyclic diguanylic acid

CC (c-di-GMP).
 CC -!- PATHWAY: Bacterial cellulose biosynthesis.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (potential).
 CC -!- DOMAIN: There are two conserved domains in the globular part of
 the catalytic subunit: the N-terminal domain (domain A) contains
 the conserved DXD motif and is possibly involved in catalysis and
 substrate binding. The C-terminal domain (domain B) contains the
 OXXRW motif and is present only in processive glycosyl
 transferases. It could be involved in the processivity function of
 the enzyme, possibly required for holding the growing glycan chain
 in the active site.

-!- INDUCTION: Cellulose is produced at a linear rate with respect to
 cell growth when O(2) is present.

-!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to licensee@isb-sib.ch).

CC

DR EMBL: M37202; AAA21884.1; - .

DR PIR: A43735; AA3735.

DR InterPro: IPR003919; Cellsynth_A.

DR InterPro: IPR001173; Glycos_transf_2.

DR Pfam: PF00535; Glycos_transf_2; 1.

DR PRINTS: PRO1439; CELLSNTHASE.

KW Cellulose biosynthesis; Transferase; Glycosyltransferase;
 KW Transmembrane; Inner membrane.

FT DOMAIN 147 240 CATALYTIC SUBDOMAIN A.

FT TRANSEME 317 377 CATALYTIC SUBDOMAIN B..

FT TRANSEME 26 46 POTENTIAL.

FT TRANSEME 47 67 POTENTIAL.

FT TRANSEME 108 128 POTENTIAL.

FT TRANSEME 167 187 POTENTIAL.

FT TRANSEME 407 427 POTENTIAL.

FT TRANSEME 430 450 POTENTIAL.

FT TRANSEME 468 488 POTENTIAL.

FT TRANSEME 516 536 POTENTIAL.

FT TRANSEME 549 569 POTENTIAL.

FT TRANSEME 589 189 POTENTIAL.

FT ACT_SITE 333 333 POTENTIAL.

FT SITE 236 236 SUBSTRATE BINDING (POTENTIAL).

FT SITE 238 238 SUBSTRATE BINDING (POTENTIAL).

SQ SEQUENCE 754 AA; 84442 MW; 8D5C1FER2E2C068 CRC64;

Query Match 4.0%; Score 228.5; DB 1; Length 754;
 Best Local Similarity 19.3%; Pred. No. 1. 2e-08;
 Matches 154; Conservative 84; Mismatches 229; Indels 329; Gaps 30;

QY 263 MRLVVLCLFLFLHYRITNPVPAFALWL-----VSVICETWFALSWILDQFQPKWPVNRE 316

Db 80 VISALVSLRYLTWRLFELD-FDTWIGGLGVTLMAELYKLYMFLSFYFOTIOPHLRA 137

Qy 317 TYLDRLAFLRYRREGESQAAVDFSTVDPLKEPPVLTANVLSLAVDXPVDKVSCYY 376

Db 138 P---LPLPDNDWPH---VDFIFPYD--EQLSIVRLVFLGAIGIDNPDPKIVNYI 186

Qy 377 FDDGAAMLSFESLAETSEFAKRWVPCFKKYKSYIEPRAPEWYFRAKIDLKDQVTSFVDR 436

Db 187 LDDG----- 190

Qy 437 RAMKREVEFKTRINALVSKALKCPSEGGWMODGTWPGNNNTGDPGMQVFLGQNGLD 496

Db 191 --VRPFEEQFAKDCGAL----- 205

Qy 497 AEGNELPLRVVSRERKPGFOHHKAGMANLVRSVLTNGPFPFLNLDCPHYINNSKL 556

Db 206 -----VIGRVD---SSHAKAGNLNAIK---RTSDYIILDCDI-IPTRAFL 246

CC

CC -!- COFACTOR: Magnesium (By similarity).

CC -!- ENZYME REGULATION: Activated by bis-(3',-5') cyclic diguanlyclic acid

CC (c-di-GMP) (By similarity).

CC -!- PATHWAY: Bacterial cellulose biosynthesis.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 (potential).

CC -!- DOMAIN: There are two conserved domains in the globular part of

QY 557 REAMCFMLDPNLGKQCVYVOFFORFDGIDKNDRYANRNTVFFDINLRGLDG1QGPVVGT 616

Db 247 QIAGMMWV---ADRKIALMOTPHFYSPDP----- 273

QY 617 GCVFNRATLYGEPPTIKVKHKKPSSLLSKLCGGSRRKNSKAKKEDKKSGRHINDSTVPVF 676

Db 274 ---FQRNLAVGRTP----- 288

QY 677 NLDDDEECVEGAGFDDKALLMSQMSLEKRGQSOSAVFASLMEENGVPPSATPENLKE 736

Db 289 NL---FVGVIQDND-----FMDATFFCGSCAI-----LRRE 317

QY 737 ATHVISCGYEDKSDWGMIGWIGSVTBDILTGFKMHARGWASVTCYCMPLPAFKGSAPIN 796

Db 318 ALESIG-----GFAVETVDAHFMQRGWSAYL--RIPASGLATER 362

QY 797 LSDRDLNQVLRWALGSEVLFSLRHCPIWYGYNCRKLFERAYVNTTIPITSIPPLMCT 856

Db 363 LTHHIGQMRMWRARGMIQI-FRDVNPW--LGGGLKGORCLCYLISAMTSFFFIPRVIFLA 418

QY 857 LLAWCLFNFQFIPPOISIASIWFSLFSIATGLEMWWS--GYGIDEMWRNEFW- 912

Db 419 SPLAFLFQQNTIAASP-----LAVLAYATPHMFHSTAATAVKNGWR-YSPWS 466

QY 913 VLGGSVAHLFALVFGQILKVLAGIDTNFVTKA--SDEDGDFAEYLFLKWMTLILPTT 969

Db 467 EVYETTMALFLRVVTITLMPSKKGKVENTGGVLEEEFDLQTY-----PNI 516

QY 970 LILTVNLVGVVAGVSYATINSYCGOSWGPFLFG--KLFKAFWVIVHLYPFLKGLMGRHQNRPTI 1027

Db 517 IF-----AGIMTGLLIGLFELTFHF-----NQLAGIAKRAYLL 550

Qy 1028 VVWWS-----VILASI 1038

Db 551 NCIWAMISLILLAI 566

OC

OC Pseudomonas fluorescens.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas

OX NCBI_TAXID=294;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SBW25;

RX MEDLINE=2203850; PubMed=12019221;

RA Spiers, A.J.; Kahn, S.G.; Bohannon, J.; Travissano, M.; Rainey, P.B.;

RT Adaptive divergence in experimental populations of Pseudomonas fluorescens. I. Genetic and phenotypic bases of wrinkly spreader fitness." Genetics 161: 33-46 (2002).

RL

CC -!- FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes uridine 5'-diphosphate glucose to cellulose, which is produced as an extracellular component responsible for the structural integrity and rigidity of self-supporting mats characteristic of the "wrinkly spreader" phenotype.

CC -!- CATALYTIC ACTIVITY: UDP-glucose + ((1,4-beta-D-glucosyl)(N) = UDP + ((1,4-beta-D-glucosyl)(N)).

CC

CC the protein; the N-terminal domain (domain A) contains
 CC the conserved DxD motif and is possibly involved in catalysis and
 CC substrate binding. The C-terminal domain (domain B) contains the
 CC OXXW motif and is present only in processive glycosyl
 CC transferases. It could be involved in the processivity function of
 CC the enzyme, possibly required for holding the growing glycan-chain
 CC in the active site.

-> SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC in the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>).
 CC or send an email to license@isb-sib.ch).

DR EMBL: AV074776; AAL17142.1;

KW Cellulose biosynthesis; Transferase; Glycosyltransferase; .

KW Transmembrane; Inner membrane.

FT DOMAIN 157 250 CATALYTIC SUBDOMAIN A.

FT DOMAIN 327 387 CATALYTIC SUBDOMAIN B.

FT TRANSMEM 36 55 POTENTIAL.

FT TRANSMEM 59 76 POTENTIAL.

FT TRANSMEM 83 101 POTENTIAL.

FT TRANSMEM 116 138 POTENTIAL.

FT TRANSMEM 417 436 POTENTIAL.

FT TRANSMEM 440 462 POTENTIAL.

FT TRANSMEM 524 546 POTENTIAL.

FT TRANSMEM 551 573 POTENTIAL.

FT ACT_SITE 199 199 POTENTIAL.

FT ACT_SITE 343 343 POTENTIAL.

FT SITE 246 246 SUBSTRATE BINDING (POTENTIAL).

FT SITE 248 248 SUBSTRATE BINDING (POTENTIAL).

SQ 739 AA: 82165 MN; 2B962EA3854B23BB CRC64;

Query Match 4.0%; Score 228; DB 1; Length 739;
 Best Local Similarity 18.8%; Pred. No. 1.3e-08; Matches 213; Gaps 33;
 Best 149; Conservative 87; Mismatches 342; Indels 342;

QY 257 PYRNVIMRLVNLCL----FLHVRITNPVNPAFALWL----VSVICETWFAL---- 300

Db 80 PGRLAI-LALWVLSSLAISRYMEFLWLTSL--GEEIWDMFFGGGLVAEFAFYALIYLIFG 136

QY 301 ---SWILDQFPKWPVNPNEYTLDRRALYDRECEGPSOLAADVADIFVSTVDPLKEPPVLA 356

Db 137 YVQTAWPPLRTRPWL-----KTEPEEMPTVDVPIPTN--EALSTIV 176

QY 357 NTVLISLLAVDPYDVKSCIVFDDGAAMLSFESLAEITSEPKARKWWPFKKYSIERRPENY 416

Db 177 LTIFAAQAMDWPKDQLRIVHWDG-----RDDRFR---FCKVGVN----- 215

QY 417 FAAKIDYLKDVKVTSFKDRRAMKREYEEFKIRINALVSKALKCPEEGWWMDGTWPWPGN 476

Db 216 ----- 215

QY 477 NTGDHPGMIQVFLGQNGGLDAEGNELPRLVYVSREREKPGFOHHKKAGAMNALVRVASLT 536

Db 216 -----YIRRN-----NFHAKAGNLNEALKV---T 237

QY 537 NGPFPLNLDCDHYINNSKALREAMCFLMDPNLGKQCVYQFQRF--DGIDKNDRY-- 590

Db 238 DGIVTALEFDADHVPTRSFLQVSLGWFLWKLPKLDG----AMQPQTHFESPFPEKFNDTFR 293

QY 591 -ANRNTVFDINLRGLDGIGQGPVYVWGTGCVFNRTIALYGEPPITKVKHKPKSPSSLRKCGS 649

Db 294 VPNEGELFYGLVQDGNDLNATFGSCAVIR-----EP----- 328

QY 650 RKKNSAKKESDKKKSGRHTDSTVVFNLDDIEEGVEGAGFDEKKLMSQMSLEKRGQ 709

Db 329 ----- 328

QY 710 SAVFVASTLMENGVPPPSATPENLKEATHVISGQGYEDKSDWGMHEIGWIQGIVSVTEDILTG 769

RL	J. Biol. Chem. 265:4782-4784(1990).	FT	TRANSMEM	106	126	POTENTIAL.
RN	[6] 3D-STRUCTURE MODELING, AND MUTAGENESIS OF D-188; D-189; D-236; D-333;	FT	TRANSMEM	398	418	POTENTIAL.
RP	O-369; R-370 AND R-372.	FT	TRANSMEM	423	443	POTENTIAL.
RC	SRAIN-ATCC 23769;	FT	TRANSMEM	468	488	POTENTIAL.
RX	PubMed=11430986;	FT	TRANSMEM	507	527	POTENTIAL.
RA	Saxena I.M.; Brown R.M. Jr.; Dandekar T.;	FT	TRANSMEM	547	567	POTENTIAL.
RT	"structure-function characterization of cellulose synthase: relationship to other glycosyltransferases.";	FT	ACT-SITE	1513	1533	POTENTIAL.
RL	Phytochemistry 57:1135-1148(2001).	FT	ACT-SITE	189	189	POTENTIAL.
[7]	REVIEW ON DOMAIN ARCHITECTURE.	FT	SITE	333	333	POTENTIAL.
RX	PubMed=883697;	FT	SITE	236	236	SUBSTRATE BINDING (POTENTIAL).
RA	Saxena I.M.; Brown R.M. Jr.; Fevre M.; Geremia R.A.; Henrissat B.;	FT	SITE	238	238	SUBSTRATE BINDING (POTENTIAL).
RT	"Multidomain architecture of beta-glycosyl transferases: implications for mechanism of action.";	FT	MUTAGEN	188	188	D->N: DECREASE IN ACTIVITY.
RT	J. Bacteriol. 177:1419-1424(1995).	FT	MUTAGEN	188	188	D-P: LOSS OF ACTIVITY.
RL	-1- FUNCTION: Bifunctional protein comprised of a catalytic subunit and a regulatory subunit. The catalytic subunit of cellulose synthase polymerizes uridine 5'-diphosphate glucose to cellulose in a processive way. The thick cellulosic mats generated by this enzyme probably provide a specialized protective environment to the bacterium. The regulatory subunit binds bis-(3',5') cyclic di-glycosylic acid (c-di-GMP).	FT	MUTAGEN	189	189	D-Y: LOSS OF ACTIVITY.
CC	-1- CATALYTIC ACTIVITY: UDP-glucose + {(1,4)-beta-D-glucosyl}(N) = UDP + [(1,4)-beta-D-glucosyl](N+).	FT	MUTAGEN	236	236	D-Y: LOSS OF ACTIVITY.
CC	-1- COFACTOR: Magnesium (By similarity).	FT	MUTAGEN	333	333	D-R: LOSS OF ACTIVITY.
CC	-1- ENZYME REGULATION: Activated by c-di-GMP.	FT	MUTAGEN	369	369	Q->M: LOSS OF ACTIVITY.
CC	-1- PATHWAY: Bacterial cellulose biosynthesis.	FT	MUTAGEN	370	370	R-P: LOSS OF ACTIVITY.
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane (Potential).	FT	MUTAGEN	370	370	R->O: DECREASE IN ACTIVITY.
CC	-1- DOMAIN: There are two conserved domains in the globular part of the catalytic subunit: the N-terminal domain (domain A) contains the conserved DxD motif and is possibly involved in catalysis and substrate binding. The C-terminal domain (domain B) contains the QXXRW motif and is present only in processive glycosyl transferases. It could be involved in the processivity function of the enzyme, possibly required for holding the growing glycan chain in the active site.	FT	MUTAGEN	372	372	R->A: LOSS OF ACTIVITY.
CC	-1- INDUCTION: Cellulose is produced at a linear rate with respect to cell growth when O(2) is present.	SO	SEQUENCE	1550 AA:	168161 MW:	63AB8952BC39E961 CRC64;
CC	-1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.	Query	Match	3.5%	Score	196; DB 1; Length 1550;
CC	-1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE FAMILY.	Matches	Best Local Similarity	20.1%	Pred. No.	8.3e-06;
CC	-1- CAUTION: Was originally (Ref.1) thought to be two separate ORFs named aceA and aceB, due to a frameshift in position 678.	Matches	161; Conservative	93;	Mismatches	241; Indels
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).	Db	Gaps	306;	Gaps	36;
CC	-----	Db	Db	126 SYFQTIAPLRAP---LPL---PPNPDEWPPTYDIFVPTN---EELSVIPLVLSIG	QY	251 PSSRNPYRNIMRLRVICLFLHYRITNPVPNAFLW-----VSVICEIMHALSWIL 304
CC	-----	Db	Db	365 VDVYVDKVSCVYEDGAAMLSFESELSEFARKWPKCKYSIEPRAPWMAAKIYL 424	QY	70 PSRRSQIF-LEVLSGLVSL-RYLWWRITEL- SFDTWLQLGLGTMLVIAELKAMMIEL 125
CC	-----	Db	Db	305 DOFPKFWPNRRETYDLRLRYDREGERSOLAAYDIFVSTVDPLKEPLPVLTSLA 364	QY	175 IDWPEEKVVRHILDG-----RREPE-FAA-----197
CC	-----	Db	Db	126 SYFQTIAPLRAP---LPL---PPNPDEWPPTYDIFVPTN---EELSVIPLVLSIG	QY	425 KDKVQTSFKDRRAMKREYEFKIRRNALVSKALKCPEEGWVMDGTTPWGNNTGDHFCM 484
CC	-----	Db	Db	198 -----FAAEC-----	QY	203 -----GAN-----VIAKPT-----NEHAKAGLNIAIG----HTDGDIYLIF 235
CC	-----	Db	Db	485 IOKFQGONGGLDAEKGELNPLRVVYSEKRGFOHHKKAGAMNALVRAVSLNGPFTNL 544	QY	545 DCDDYINNSKALREAMCFLM-DPNULGKQVYVFQRFQRTGIDKNDRYANRNFVDINLR 603
CC	-----	Db	Db	236 DCDH-VPTRAFLQTMGMWAEVDP---KIALMOPPHFYSPDP-----273	QY	604 GLDGIQGPVYVGICCVFNRTALGYEPPIKVHKHKPSLLSKLOGGSRKNSKAKKESDK 663
CC	-----	Db	Db	274 -----FORNLSAGYRTP-----285	QY	664 KSGRHIDSTPVFMUDDLEFEGVEGAGDFEKKALIMQSLEKRRGQSAVFIESTMENG 723
CC	-----	Db	Db	285 -----PEGNL--PYGVWQDGND-----FWDATFFCGSCAI---313	QY	724 VPPSATPENILKELIIVHISCGYEDBKSDGMEMIGWIYGSVTEIDLTFGKMHARWRSIYM 783
DR	EMBL: X54676; CAA38487; 1; ALT_FRAME.	Db	Db	314 -----LRRTAEIQIG-----GRATQVTEADAHALKMQRJGWSTAYL- 350	QY	784 PKLPAFKGSAPINUSDRNLNQVLRLAHLGSEVILSRHCPIWYGNGR-LKFLEFAYVWT 842
DR	EMBL: X54676; CAA38488; 1; ALT_FRAME.	Db	Db	351 -RIPLLAGGLATERLILTHIGORVWRAGLMOI-FRIDNPLF---GRGLSWGRDLCYLISAM 404	QY	902 ---IDEWWNEQEW-VIGGSVAHFIAVQFGILKVLVLAGIDTNFTVTSKASDEGDFAEYL 957
DR	PFAM: PF00555; Glycos_transf_2; 1.	Db	Db	453 ASKINKGWR-YSFWSSEVYETTMALFLVRVITVLLSPSRGKENVTDKGGLLERGYFDL- 509	QY	902 ---IDEWWNEQEW-VIGGSVAHFIAVQFGILKVLVLAGIDTNFTVTSKASDEGDFAEYL 957
KW	Cellulose biosynthesis; Transferase; Glycosyltransferase; Transmembrane; Inner membrane.	QY	QY	843 IYPFTSIPILMYCILAVLCLFLNQFLPQISNIAISWFLSFSIFATGILEURWSVG- 901	QY	405 TSFLFAVPRVIFLSSPLAFLFFGONIASSP-----LALLAIAPM-HFAVGT 452
DR	InterPro: IPR001173; Glycos_transf_2.	QY	QY	902 ---IDEWWNEQEW-VIGGSVAHFIAVQFGILKVLVLAGIDTNFTVTSKASDEGDFAEYL 957	QY	902 ---IDEWWNEQEW-VIGGSVAHFIAVQFGILKVLVLAGIDTNFTVTSKASDEGDFAEYL 957
DR	PRINTS: PR01439; CELLSWVHASBA.	Db	Db	453 ASKINKGWR-YSFWSSEVYETTMALFLVRVITVLLSPSRGKENVTDKGGLLERGYFDL- 509	QY	902 ---IDEWWNEQEW-VIGGSVAHFIAVQFGILKVLVLAGIDTNFTVTSKASDEGDFAEYL 957
DR	TRANSMEM	FT	TRANSMEM	106	126	POTENTIAL.
FT	TRANSMEM	FT	TRANSMEM	398	418	POTENTIAL.
FT	TRANSMEM	FT	TRANSMEM	423	443	POTENTIAL.
FT	TRANSMEM	FT	TRANSMEM	468	488	POTENTIAL.
FT	TRANSMEM	FT	TRANSMEM	507	527	POTENTIAL.
FT	TRANSMEM	FT	TRANSMEM	547	567	POTENTIAL.
FT	TRANSMEM	FT	TRANSMEM	1513	1533	POTENTIAL.
FT	TRANSMEM	FT	TRANSMEM	189	189	POTENTIAL.
FT	ACT-SITE	FT	ACT-SITE	333	333	POTENTIAL.
FT	ACT-SITE	FT	ACT-SITE	236	236	SUBSTRATE BINDING (POTENTIAL).
FT	ACT-SITE	FT	ACT-SITE	238	238	SUBSTRATE BINDING (POTENTIAL).
FT	ACT-SITE	FT	ACT-SITE	188	188	D->N: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	188	188	D-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	189	189	D-Y: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	236	236	D-Y: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	333	333	D-R: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	369	369	Q->M: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	370	370	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->A: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R-P: LOSS OF ACTIVITY.
FT	ACT-SITE	FT	ACT-SITE	372	372	R->O: DECREASE IN

RESULT 14

ID COX1_PARLI	STANDARD;	PRT;	517 AA.
AC P12700;			
DT 01-OCT-1989 (Rel. 12, Last sequence update)			
DT 15-JUN-2002 (Rel. 41, Last annotation update)			
DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).			
COI.			
OS Paracentrotus lividus (Common sea urchin).			
OG Mitochondrion.			
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoidea; Echinidae; Paracentrotus; OX NCBI_TAXID=7655;			
RP [1]			
SEQUENCE FROM N.A.			
MEDLINE-#9291831; PubMed=254576;			
Cantatore P., Roberti M., Rainaldi G., Gadaleta M.N., Saccone C.; RT "The complete nucleotide sequence, gene organization, and genetic code of the mitochondrial genome of Paracentrotus lividus.";			
J. Biol. Chem. 264:10965-10975(1989). [2]			
RP SEQUENCE OF 469-517 FROM N.A.			
RT TISSUE-EGG; RX MEDLINE-#87248108; PubMed=3596250;			
RA Cantatore P., Roberti M., Morisco P., Rainaldi G., Gadaleta M.N., Saccone C.; RT "A novel gene order in the Paracentrotus lividus mitochondrial genome."; RL Gene 53:41-54(1987).			
-1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3 AND COPPER B.			
CC -1- CATALYTIC ACTIVITY: 4 ferrocyanochrome c + O ₂) = 4 ferricyanochrome c + 2 H ₂ O.			
CC -1- PATHWAY: Respiratory chain; terminal step.			
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL.			
CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.			
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.			
CC			
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See http://www.isb-sib.ch/announce/ or send an email to license@sb-sib.ch).			
CC -----			

RESULT 15

ID YKTA_CAEEL	STANDARD;	PRT;	590 AA.
AC P34322;			
DT 01-FEB-1994 (Rel. 28, Last sequence update)			
DT 15-JUN-2002 (Rel. 41, Last annotation update)			
DE Hypothetical protein C07A9_11 in chromosome III.			
GN C07A9_11.			
OS Caenorhabditis elegans.			
OC Cukarvota; Metazoa; Nematoidea; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderaida; Caenorhabditis.			
OX NCBI_TAXID=6239;			
RP [1]			
SEQUENCE FROM N.A.			
RC STRAIN-BRISTOL N2;			
RX MEDLINE-#91150718; PubMed=7906398;			
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., EMBL; M16519; AAA31989.2; .			
DR PIR; C34284; C34284.			
DR HSSP; P00396; 20C.			
DR InterPro; IPR000883; COX1.			
DR Pfam; PF00115; COXI; 1.			
DR PRINTS; PR01165; CYCOXIDASE1.			
DR PROSITE; PS00077; COXI; 1.			
KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane; Respiratory chain; Inner membrane.			

RT RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RL Wohlwend P.;
 RT elegans.";
 CC Nature 368:32-38 (1994).
 CC
 CC -|- SIMILARITY: SOME, TO YEAST YJRR106w.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; 229094; CA82341.1; -.
 DR PIR; SA0707; S40707.
 DR WormPep; C07A9_11; CE00501.
 DR InterPro; IPR004837; Naca_Exmemb.
 DR Pfam; PF01699; Na,Ca_Ex; 2.
 KW Hypothetical protein.
 SQ SEQUENCE 590 AA; 65540 MW; 71ADE3945C311BCB CRC64;
 Query Match 2.1%; Score 118.5; DB 1; Length 590;
 Best Local Similarity 19.3%; Pred. No. 0.84; Gaps 26;
 Matches 111; Conservative 89; Mismatches 205; Indels 171; Gaps 26;
 QY 531 VSAVLTNGPFLNLDCDHYINNSKALREAMCFMLMDPNLGKOVQVFQPFORPDGIDKNDRY 590
 DB 97 VSTLITSPFDVEV-----FSTIIDLFLVVALSF-LAFCV-----F 133
 OY 591 ANRNTVFFDINLRGLDG1QGPVYVGTCGVFNRTALYGYEPPIKVKKPSULLSKLCGGSR 650
 DB 134 YNRVTLWMLTFIQLLILVITVIGQAQAVHNR -----KTKAL-----Q 171
 QY 651 KRNNS-KAKKEDKKSGRHTDSTPVFVNLDIFEGVEGAEF-----DEDEKALMS 699
 DB 172 KONSTSRSKSIKSLRSRKSIHSVAPPMPVPEIEVHDQEAFFPEISVUVGAIDKLKEHMAE 231
 QY 700 QMSLEKRGQSAFVASTLMEENGVPPSATPENLKLKEATHVISCGYEDSDWGMEIGWY 759
 DB 232 KAQTTRTRYTKRASFMVNGDGNLNGLHPYATHNL-----GIRRSELSDEBEEFWVH 285
 QY 760 GSV-----TEDLIGFKMH-----ARGWTSIYCMPKLPF 789
 DB 286 GHVFQHGHEARSRAASLVLVFPMQIKWSRSKDIKLDLAELHDPRBAEDWEEMNIFSKVMAY 345
 QY 790 KGSSAPPINLSDRLNQVLRWALGSVEILFLSRHCPITWGYNGRLKLERFAVNTTLYPISSI 849
 DB 346 INTVP-----NILFKLTIPLNEMWSK-----PLT-- 370
 QY 850 PLIMYCTLLAVCLEFTNQFII-PQISNIASIWIFLSFLSIFATGLEMRSVGVDEWRN 908
 DB 371 -LLHWTGCPAFLFLSIOFLETPFGSPGLWVYGLAVSI-VLAAILIVMTEYLSPQVKYK 428
 QY 909 EQTWVIGGS---AHLFAYFOGIKVLAGIDTNFTVTSKASDEODGDAEFLYLFKWTLLIP 966
 DB 429 EIVSYSGFIMSIWIYLISSEVVNVV---TMLGWVRSVSH-----VIGITILAWS----- 476
 QY 967 PTFLLILVNUG-WVAGVSYAINTSGYQSW-----GFLFGKLFPAFWVIVHLYPF---L 1014
 DB 477 -----NSGDLIADVS-WVKOGYPRMAMAAIGGPLF-NLMGFGL-----PTIAKL 522
 QY 1015 KGLMGQRQRPTI-WVWSVLLASIFSLWVRLDPF 1049
 DB 523 QGKVISMINTPYRLILFLAISILATLIGIPVOKF 558

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 11:04:03 : Search time 49 Seconds
(Without alignments)
4478.371 Million cell updates/sec

Title: AAC39336
Perfect score: 5677
Sequence: MEESEGETAGKPKMKNIVPOTC.....IDPFITSRVTGPDILEGGC 1065

Scoring table: BLOSSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPREMBL_21:
1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_micr:
8: sp_organelle:
9: sp_phage:
10: sp_plant:
11: sp_rabbit:
12: sp_virus:
13: sp_vertebrate:
14: sp_unclassified:
15: sp_rvirus:
16: sp_bacteriopl:
17: sp_archeap:

Pred No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	5677	100	0	048948
2	5654	99	10	048948
3	4954	87	3	09XGK6
4	4571	80	5	09LL15
5	4564	80	4	09LL11
6	4529	79	10	09LL15
7	4028	71	0	1075
8	4024	70	9	1074
9	3984	70	9	1081
10	3900	68	7	1059
11	3875	68	3	1086
12	3843	67	7	1094
13	3811	67	1	1026
14	3805	67	0	1026
15	3801	67	0	1065
16	3741	65	9	1043

SEQUENCE FROM N.A.
RC STRAIN=CV; COLUMBA;
RX MEDLINE=98111412; PubMed=9445479;

RA Arioli T., Peng L., Bettner A.S., Burn J., Wittke W., Herth W.,
RA Camilleri C., Hofte H., Plazinski J., Birch R., Cork A., Glover J.,
RA Redmond J., Williamson R.E.;

RT "Molecular analysis of cellulose biosynthesis in Arabidopsis.";
RL Science 279:717-720(1998);
DR EMBL: AF027114; AAC39336; 1. -

DR Q9fkh6 arabidopsis
DR InterPro: IPR005180; Cellulose_synt.

DR Pfnam: PR03552; Cellulose_synt; 1.

DR SMART: S00014; RING; 1.

SEQUENCE 1065 AA; 119659 MW; 1EF01C95FB05BEAC CRC64;

Query Match 100.0%; Score 5677; DB 10; Length 1065;
Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;
Matches 1065; Conservative 0;

QY 1 MEESEGETAGKPKMKNIVPOTC.....IDPFITSRVTGPDILEGGC 1065;
Db 1 MESEGETAGKPKMKNIVPOTC.....IDPFITSRVTGPDILEGGC 1065;
QY 61 SPQCKTRKRKKSPAIRPODKDEGLADEGTVEINYPOKEK1SRMLGMHHLTRKGEM 120

ALIGNMENTS

RESULTS

RESULT	ID	PRELIMINARY	PRT	1065 AA.
048948	048948;	048948;	DT	01-JUN-1998 (TREMBLrel. 06, Created)
			DT	01-JUN-1998 (TREMBLrel. 06, Last sequence update)
			DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)
			DE	Cellulose synthase catalytic subunit.
			GN	ATH-B.
			OS	Arabidopsis thaliana (Mouse-ear cress).
			OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
			OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
			OC	eurosid s II; Brassicales; Brassicaceae; Arabidopsis.
			OX	NCBL_TaxID=3702;
			RN	[1]
			RP	SEQUENCE FROM N.A.
			RC	STRAIN=CV; COLUMBA;
			RX	MEDLINE=98111412; PubMed=9445479;
			RA	Arioli T., Peng L., Bettner A.S., Burn J., Wittke W., Herth W., RA Camilleri C., Hofte H., Plazinski J., Birch R., Cork A., Glover J., RA Redmond J., Williamson R.E.;
			RT	"Molecular analysis of cellulose biosynthesis in Arabidopsis.";
			RL	Science 279:717-720(1998);
			DR	EMBL: AF027114; AAC39336; 1. -
			DR	InterPro: IPR005180; Cellulose_synt.
			DR	PR03552; Cellulose_synt; 1.
			DR	SMART: S00014; RING; 1.
			SEQUENCE	1065 AA; 119659 MW; 1EF01C95FB05BEAC CRC64;

Query Match 100.0%; Score 5677; DB 10; Length 1065;
Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;
Matches 1065; Conservative 0;

QY 1 MEESEGETAGKPKMKNIVPOTC.....IDPFITSRVTGPDILEGGC 1065;
Db 1 MESEGETAGKPKMKNIVPOTC.....IDPFITSRVTGPDILEGGC 1065;
QY 61 SPQCKTRKRKKSPAIRPODKDEGLADEGTVEINYPOKEK1SRMLGMHHLTRKGEM 120

Db	61	SCPOCKTRKRLKGSPAIFGDKDGLADEGVENFNPQEKEISERMGWHLTRGKEM	DT	01-MAR-2001 (TREMBlrel. 16, Last sequence update)
Oy	121	GEOYDKEVSHNHLPLRITSRQDTSGEFSASASPERLSVSYSTIAGKRKLYPSSDNQSNRR	DE	01-JUN-2002 (TREMBlrel. 21, Last annotation update)
Db	121	GERQYDKEVSHNHLPLRITSRQDTSGEFSASASPERLSVSYSTIAGKRKLYPSSDNQSNRR	OS	Cellulose synthase catalytic subunit.
Oy	181	IYPVGIGLGNVANKERVDGKMKOEKNTPQSYTOAASERGGVVIDASDILADELLNDEA	OC	Arabidopsis thaliana (Mouse-ear cress).
Db	181	IYPVGIGLGNVANKERVDGKMKOEKNTPQSYTOAASERGGVVIDASDILADELLNDEA	OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Oy	241	ROPLSRKVSISSRINPRMVIMRLVILCFLHYRTNPVPNAFAWMVLVSICEWFAL	OC	Spermatophyta; Magnoliophyta; eudicots; Rosidae;
Db	241	ROPLSRKVSISSRINPRMVIMRLVILCFLHYRTNPVPNAFAWMVLVSICEWFAL	OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.
RN	{1}		NCBI_TaxID:3702;	
RP			SEQUENCE FROM N.A.	
RC			SPRAIN-COLUMBIA;	
RX			MEDLINE:9-9397451; PubMed=10470850;	
RA			Kaneko T., Katoh T., Saito S., Nakamura Y., Asamizu E., Kotani H.,	
RA			Miyajima N., Tabata S.:	
RT			"Structural analysis of Arabidopsis thaliana chromosome 5. IX. Sequence features of the regions of 1,011,550 bp covered by seventeen	
RT			P1 and TAC clones."	
RT			DNA Res. 6:183-195(1999).	
RL			EMBL: AB018111; BAB09693.1; -.	
DR			DR InterPro; IPR005150; Cellulose_synt.	
DR			DR InterPro; IPR001841; ZnfRing.	
DR			PFam: PF03522; Cellulose_synt; 1.	
DR			SMART: SM00184; RING: 1; PEPPIANTIVL	
SO	SEQUENCE	1065 AA; 119682 MW; 3AA4714C53C4D581 CRC64;		
Db	421	IDYLKDQVTSFKDRRAMREVEEKIRINALVSKALKCPERGWMDGTWPGNNITGD	480	
Oy	361	SILAVDYPDKVSYVDDGAMLSFEISAESEFARKWPKCKYSIEPRAPEWYFAK	420	
Db	361	SILAVDYPDKVSYVDDGAMLSFEISAESEFARKWPKCKYSIEPRAPEWYFAK	420	
Oy	421	IDYLKDQVTSFKDRRAMREVEEKIRINALVSKALKCPERGWMDGTWPGNNITGD	360	
Db	421	IDYLKDQVTSFKDRRAMREVEEKIRINALVSKALKCPERGWMDGTWPGNNITGD	480	
Oy	481	HGMIOVFLGONGGLDAEGNEPLRVTYVSRERKPGFOHKKAGAMNALVRVSAVLNGPF	540	
Db	481	HGMIOVFLGONGGLDAEGNEPLRVTYVSRERKPGFOHKKAGAMNALVRVSAVLNGPF	540	
Oy	541	ILNLDCDHYINNSKALRACMFCLMDPNLGKQCYWQFQREFGKIDNDYRANNTVFDI	600	
Db	541	ILNLDCDHYINNSKALRACMFCLMDPNLGKQCYWQFQREFGKIDNDYRANNTVFDI	600	
Oy	601	NLRGLDGQGPVYVGTCVFNRTALLYGEPPIKVKHKPSLISKLCGGSRKNSKAKKES	660	
Db	601	NLRGLDGQGPVYVGTCVFNRTALLYGEPPIKVKHKPSLISKLCGGSRKNSKAKKES	660	
Oy	601	NLRGLDGQGPVYVGTCVFNRTALLYGEPPIKVKHKPSLISKLCGGSRKNSKAKKES	660	
Db	601	NLRGLDGQGPVYVGTCVFNRTALLYGEPPIKVKHKPSLISKLCGGSRKNSKAKKES	660	
Oy	661	DKKSGSRHTSDSTVPUFNLDIEBZVEGAGFDEKKALMSOMSLEKRGQSAFWASTIME	720	
Db	661	DKKSGSRHTSDSTVPUFNLDIEBZVEGAGFDEKKALMSOMSLEKRGQSAFWASTIME	720	
Oy	721	NGVPPSATPENLKEATHVLSGYEDKSDWGMELIGWIGWIGSVTEDIQFGKMHARGRSI	780	
Db	721	NGVPPSATPENLKEATHVLSGYEDKSDWGMELIGWIGWIGSVTEDIQFGKMHARGRSI	780	
Oy	781	YCMPLKPAFKGSAPINSDRMLNQYRMLAGSWEILSFHCP1IWGYNQRKLERFIYNN	840	
Db	781	YCMPLKPAFKGSAPINSDRMLNQYRMLAGSWEILSFHCP1IWGYNQRKLERFIYNN	840	
Oy	841	TIIPTISIPLIMCTIILAVCLFTNOTIPQSNIASTWFLSLSPATGIELMRNSV	900	
Db	841	TIIPTISIPLIMCTIILAVCLFTNOTIPQSNIASTWFLSLSPATGIELMRNSV	900	
Oy	841	TIIPTISIPLIMCTIILAVCLFTNOTIPQSNIASTWFLSLSPATGIELMRNSV	900	
Db	901	GIDEWWRNEQEWVYIGGSAHLPAVFOGTLKLVLAGIDNFNTVSKASDEDGDFAEPLYFW	960	
Oy	901	GIDEWWRNEQEWVYIGGSAHLPAVFOGTLKLVLAGIDNFNTVSKASDEDGDFAEPLYFW	960	
Db	961	TUILLPTTILTIVNLAGVWAGVSYATINSYGQSWGPIGLKLFCAFWMVWHLVYPLKGLMR	1020	
Oy	961	TUILLPTTILTIVNLAGVWAGVSYATINSYGQSWGPIGLKLFCAFWMVWHLVYPLKGLMR	1020	
Db	1021	QNRPTTVWWSVLLASFLSLWWRDFTSRVTPGDLLEGINC 1065	1065	
Oy	1021	QNRPTTVWWSVLLASFLSLWWRDFTSRVTPGDLLEGINC 1065	1065	
Db	1021	QNRPTTVWWSVLLASFLSLWWRDFTSRVTPGDLLEGINC 1065	1065	
RESULT	2			
O9FHK6	1D	PRELIMINARY;	PRT;	1065 AA.
O9FHK6	AC			
DT	01-MAR-2001 (TREMBlrel. 16, Created)			

Oy	661	DRKSGRHTDSTPVFNUDDIEGEGAGFDDEKALLMSOMSLERKGOSAVFASTIME	720	QY	61	SCPOCKTIRKLRKGSPATPGKDDEGLADEGVETE---NYPOKEITSERMGLWHLTRGK	116
Db	661	DKKSGRHTDSTPVFNUDDIEGEGAGFDDEKALLMSOMSLERKGOSAVFASTIME	720	Db	61	SOPCKTIRKLRKGSPATPGKDDEGLADEGVETE---NYPOKEITSERMGLWHLTRGK	120
Qy	721	NGGVPSATPENILKEATHVISCGYEDKSDWMGEIGWIGYGSVTEIDLQFKMHARGWRST	780	Qy	117	GEEMGEPQDYKDVSHNLPRTSRQDSGETSAASPERLSVST-TAGKRPYSSVHQ	175
Db	721	NGGVPSATPENILKEATHVISCGYEDKSDWMGEIGWIGYGSVTEIDLQFKMHARGWRST	780	Db	121	GDBVGAFTYDKEIENHNIPLATSGOEVSGLAASPERLMSAPGVAGK-----	170
Qy	781	YCPKLPKFKGSPAINFSDRNQYRLWALGSVELLSFRCPINWGYNSLKLERFAYVN	840	Qy	176	SPNRRTDPV----GLGNWAKERVGWKMHQEKNGTPWST-QAASERRGGVIDASTD	228
Db	781	YCPKLPKFKGSPAINFSDRNQYRLWALGSVELLSFRCPINWGYNSLKLERFAYVN	840	Db	171	-SSRVVDPVREFGSSLGNVANKERVGWKMHQEKNGTPWST-QAASERRGGVIDASTD	229
Qy	841	TRYPTISIPLMCTLLAVCLFTNFIPOISNIASTIFLSELISIFTGILEMRWGSV	900	Qy	229	TLADEALLNDEARQPLSRKVSIPSSRNPYRAVIMRLVILCFLHYRTINPVNAFLW	288
Db	841	TRYPTISIPLMCTLLAVCLFTNFIPOISNIASTIFLSELISIFTGILEMRWGSV	900	Db	230	VLYDDSLNDEARQPLSRKVSIPSSRNPYRAVIMRLVILCFLHYRTINPVNAFLW	289
Qy	901	GIDEWWRNEQFWIWIGVSAHLFAVFGQILKVLAGIDTNFTVTKSADEGDFDAEYLFW	960	Qy	289	IVSVICIWALSWLQDFPKMFVPNRYTDLRALLYDREBPSQAVDFVSTDPL	348
Db	901	GIDEWWRNEQFWIWIGVSAHLFAVFGQILKVLAGIDTNFTVTKSADEGDFDAEYLFW	960	Db	350	KERPVLYANTVSLAYDYPVKYCIVSDGAAMLTFEALSETSEFARKWVPECKN	409
Qy	961	TRLIPIPTLTVLNLYGTVAGVSAYAINGYSQSMGPFLGFLFAWVWVHLYPELGKLMGR	1020	Qy	409	ERPAPENYFAKIDYLDKVOTSFVKDRAMREKEYEERKIRINALVSKALCPEEGWMO	468
Db	961	TRLIPIPTLTVLNLYGTVAGVSAYAINGYSQSMGPFLGFLFAWVWVHLYPELGKLMGR	1020	Db	410	EPRAPENYFAKIDYLDKVOTSFVKDRAMREKEYEERKIRINALVSKALCPEEGWMO	469
Qy	961	TRLIPIPTLTVLNLYGTVAGVSAYAINGYSQSMGPFLGFLFAWVWVHLYPELGKLMGR	1020	Qy	469	DCTPWPNNTDHPGMQVFLGONGGDAEGNELPRLVYSEKRGFQHKKAGAMNAL	528
Db	961	TRLIPIPTLTVLNLYGTVAGVSAYAINGYSQSMGPFLGFLFAWVWVHLYPELGKLMGR	1020	Db	470	DCTPWPNNTDHPGMQVFLGONGGDAEGNELPRLVYSEKRGFQHKKAGAMNAL	529
Qy	01-NOV-1999	(TREMBLrel. 12, Created)	Qy	529	VRVASVLTNGPTLNQCDHYTNNSKALREACFLMDPDLNGLQCVYQFPQRDGIDND	588	
Db	01-NOV-1999	(TREMBLrel. 13, Last sequence update)	Db	530	VRVASVLTNGPTLNQCDHYTNNSKALREACFLMDPDLNGLQCVYQFPQRDGIDND	589	
Db	01-MAY-2000	(TREMBLrel. 21, Last annotation update)	Db	589	RYANRNTVFEEDNLRGLDGIQPVYTGCVNRTALGYERPIKVKHKPSLISKCGG	648	
Db	01-JUN-2002	(TREMBLrel. 21, Last annotation update)	Db	590	YANRNTVFEEDNLRGLDGIQPVYTGCVNRTALGYERPIKVKHKPSLISKCGG	649	
Db	Cellulose synthase catalytic subunit.	GN	Db	649	SRKKNSK-AKESDKKKSGRHDSTPVFNUDDIEGEGAGDEKALLMSOMSLERKG	707	
Db	CELA3.	OS	Db	650	SRKKSSKSKGDKKKSGKHKDSTPVFNUDDIEGEGAGDEKALLMSOMSLERKG	709	
Db	Buccula; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	OC	Db	708	60SAVEFASTLMEENGGVPPSATPENILKEATHVISCGYEDKSDWMGEIGWIGS	767	
Db	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;	OC	Db	710	60SAVEFASTLMEENGGVPPSATPENILKEATHVISCGYEDKSDWMGEIGWIGS	769	
Db	eurosidis II; Malvales; Malvaceae; Gossypium.	OX	Db	768	TGFKMHARGWISIYCMKLPKPGSAINLSRNLNQYRLWALGSVELLSFRCPINWGY	827	
Db	NCBI_TAXID:3635;	OX	Db	770	TGFKMHARGWISIYCMKLPKPGSAINLSRNLNQYRLWALGSVELLSFRCPINWGY	829	
Db	[1]	RN	Db	828	GLKFLERFAYNTTIPITSIPLMCTLLAVCLFTNFIPOISNIASTIFLSELISI	887	
Db	SEQUENCE FROM N_A.	RC	Db	830	GRKWLERFAYNTTIPITSIPLMCTLLAVCLFTNFIPOISNIASTIFLSELISI	889	
Db	STRAIN-CV: TEXAS MARKER-1; TISSUE-COTTON FIBER;	RC	Db	888	FATGILEMRWGSVYIDEMWRNEQFWIWIGVSAHLFAVFGQILKVLAGIDTNFTV	947	
Db	MDBLINE=20058737; Published=20061223;	RA	Db	949	TKSAD	949	
Db	Laosinchai W., Cui X., Brown R.M. Jr.;	RT	Db	949	FRGILKMKWNGVQDOWMRNEQFWIWIGVSAHLFAVFGQILKVLAGIDTNFTV	949	
Db	"A full length cDNA of cotton cellulose synthase has high homology	RT	Db	949	TKSAD	949	
Db	with the Arabidopsis RSW1 gene and the cotton CelA1 gene (Accession	RT	Db	949	FRGILKMKWNGVQDOWMRNEQFWIWIGVSAHLFAVFGQILKVLAGIDTNFTV	949	
Db	No. AP-200453) (PCR 00-002); "	RT	Db	949	TKSAD	949	
Db	Plant Physiol. 122:291-291(2000).	RL	Db	949	FRGILKMKWNGVQDOWMRNEQFWIWIGVSAHLFAVFGQILKVLAGIDTNFTV	949	
Db	[2]	RN	Db	949	TKSAD	949	
Db	SEQUENCE FROM N_A.	RP	Db	949	FRGILKMKWNGVQDOWMRNEQFWIWIGVSAHLFAVFGQILKVLAGIDTNFTV	949	
Db	STRAIN-CV: TEXAS MARKER-1; TISSUE-COTTON FIBER;	RC	Db	949	TKSAD	949	
Db	ImmunoGold Labeling of Rosette Terminal Cellulose Synthesizing	RT	Db	949	FRGILKMKWNGVQDOWMRNEQFWIWIGVSAHLFAVFGQILKVLAGIDTNFTV	949	
Db	Complexes in a Vascular Plant (<i>Vigna angularis</i>).";	RT	Db	949	TKSAD	949	
Db	Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.	RL	Db	949	FRGILKMKWNGVQDOWMRNEQFWIWIGVSAHLFAVFGQILKVLAGIDTNFTV	949	
Db	InterPro: IPR05150; Cellulose_synt.	DR	Db	949	TKSAD	949	
Db	InterPro: IPR010841; Znf_f-ring.	DR	Db	949	FRGILKMKWNGVQDOWMRNEQFWIWIGVSAHLFAVFGQILKVLAGIDTNFTV	949	
Db	Pfam: PF03552; cellulose_synt; 1.	DR	Db	949	TKSAD	949	
Db	SMART; SM00184; RING; 1; MW: 0BBA2ED00590F29C CRC64;	DR	Db	949	FRGILKMKWNGVQDOWMRNEQFWIWIGVSAHLFAVFGQILKVLAGIDTNFTV	949	
Db	SEQUENCE FROM N_A.	SQ	Db	949	TKSAD	949	
Query Match	87.3%	Score 4954; DB 10; Length 1067;	Qy	950	EDGDAEFLYLMFKWTLIPLPFLLNLGVVAGISVINSQWGLR	1009	
Best Local Similarity	85.3%	PreD. No. 0;	Qy	1008	WLYPELGLMGQRNQRPITVWWSILLASIFSLWRIDPPTSRVPGDILEGINC	1065	
Matches	919;	Conservative 75; Mismatches 60; Tuples 24; Gaps 6;	Qy	1010	IHYPELGLMGQRNQRPITVWWSILLASIFSLWRIDPPTSRVPGDILEGINC	1065	

ID	Q9L15	PRELIMINARY;	PRT;	1075 AA.
QY	GEEMGEPYDK-----EVSHNHLPLRTSRRQTSGEFSAASPERLSVSTIAGK	165	AC	Q9L15;
Db	:: :	01-OCT-2000 (TEmBLrel. 15, Created)	DT	OCT 1- OCT 2000
Db	SQDVGRPKYDSEBIGLTKYDSEBIPRGYIPSNTSQ-1SGEIPGASPDHHMSPTGNIGR	179	DT	01-OCT-2000 (TEmBLrel. 15, last sequence update)
QY	RIPYS-SDVGNSPFRNRTRIVDPVGIGNAVKERVGWWMMQKEKNPGV--STQASERGV	221	DE	01-JUN-2002 (TEmBLrel. 21, last annotation update)
Db	-DIDASDILADEALNDEAORPLSKRVSIPSSRINPRYPMVMLRLVLCLEHYRITNP	280	Cellulose synthase 5.	
QY	180 RAPFPYHNHSNSPREFSGSV-GNVAWKERVGWWMMQKDKGKIPMTNGTISAPSEGKV	237	QY	zea mays (Maize).
Db	222 -DIDASDILADEALNDEAORPLSKRVSIPSSRINPRYPMVMLRLVLCLEHYRITNP	280	Db	DIDASDILADEALNDEAORPLSKRVSIPSSRINPRYPMVMLRLVLCLEHYRITNP
QY	238 GDIDASDYNMEDALLNDETROPLSKRVSIPSSRINPRYPMVMLRLVLCLEHYRITNP	297	QY	GDIDASDYNMEDALLNDETROPLSKRVSIPSSRINPRYPMVMLRLVLCLEHYRITNP
Db	281 VPNAFAFLWLVSVCEWFALSMILDQPKWPVNRETYLDRLARYDREGEPSQLAAVDI	340	QY	281 VPNAFAFLWLVSVCEWFALSMILDQPKWPVNRETYLDRLARYDREGEPSQLAAVDI
Db	: :	340	Db	: :
QY	298 VRNAYPLWLVSVICEIWALSFWSLDFQPKWPVNRETYLDRLARYDREGEPSQLAAVDI	357	QY	298 VRNAYPLWLVSVICEIWALSFWSLDFQPKWPVNRETYLDRLARYDREGEPSQLAAVDI
Db	341 FVSTVDLKEPPLTANTVNLSLAVDYPVKSCYVFDDGAAMLSFESLAETSEFARKWV	400	QY	341 FVSTVDLKEPPLTANTVNLSLAVDYPVKSCYVFDDGAAMLSFESLAETSEFARKWV
Db	358 FVSTVDLKEPPLTANTVNLSLAVDYPVKSCYVFDDGAAMLSFESLAETSEFARKWV	417	Db	358 FVSTVDLKEPPLTANTVNLSLAVDYPVKSCYVFDDGAAMLSFESLAETSEFARKWV
QY	401 PCKKYSIEPRAPEWYPAKIDYLKDQVTSFVKDRRAMKREVEEFKIRNALVSKALKC	460	QY	401 PCKKYSIEPRAPEWYPAKIDYLKDQVTSFVKDRRAMKREVEEFKIRNALVSKALKC
Db	418 PFKVYKIEPRAPEWYPSQKIDLKDHPSFKVRKDRAMKREVEEFKIRNALVSKALKC	477	Db	418 PFKVYKIEPRAPEWYPSQKIDLKDHPSFKVRKDRAMKREVEEFKIRNALVSKALKC
QY	461 PEEGWWMDGTPWPGNNTGHDHOMIQVFLGONGLDAEGNEELPRLVYVREKRGFOHK	520	QY	461 PEEGWWMDGTPWPGNNTGHDHOMIQVFLGONGLDAEGNEELPRLVYVREKRGFOHK
Db	: :	520	Db	: :
QY	478 PEEGWWMDGTPWPGNNTRHDHGMQVFLGHSGGLDEGENELPRLVYVREKRGFOHK	537	QY	478 PEEGWWMDGTPWPGNNTRHDHGMQVFLGHSGGLDEGENELPRLVYVREKRGFOHK
Db	521 KAGAMNALVRVASVLTNGPFLINLDCDHYINNSKALRABAMCFLMDPNLGKQVCYVQFQR	580	QY	521 KAGAMNALVRVASVLTNGPFLINLDCDHYINNSKALRABAMCFLMDPNLGKQVCYVQFQR
Db	538 KAGAMNALVRVASVLTNGQIMLNLDCHYINNSKALRABAMCFLMDPNLGKQVCYVQFQR	597	QY	538 KAGAMNALVRVASVLTNGQIMLNLDCHYINNSKALRABAMCFLMDPNLGKQVCYVQFQR
QY	581 FDGIDKNDRYANTNTVFDINJNLGLDGQGPVYVGTGCVENRITALGYEPPKVKHKIPS	640	QY	581 FDGIDKNDRYANTNTVFDINJNLGLDGQGPVYVGTGCVENRITALGYEPPKVKHKIPS
Db	598 FDGIDKNDRYANTNTVFDINJNLGLDGQGPVYVGTGCVENRITALGYEPPKVKHKIPS	655	Db	598 FDGIDKNDRYANTNTVFDINJNLGLDGQGPVYVGTGCVENRITALGYEPPKVKHKIPS
QY	641 LLSKLCGGSRKKNSKAKKEDKKSGRHTDSTVFPVNLDIESCGVEAGFDDERKALLMSQ	700	QY	641 LLSKLCGGSRKKNSKAKKEDKKSGRHTDSTVFPVNLDIESCGVEAGFDDERKALLMSQ
Db	: :	700	Db	: :
QY	656 FLSLSCGG-RKGSKSKRGSRKSKQHVDSVPPNFLEDIEGGEGAGFDDERKALLMSQ	714	QY	656 FLSLSCGG-RKGSKSKRGSRKSKQHVDSVPPNFLEDIEGGEGAGFDDERKALLMSQ
Db	701 MSLEKRIGQSAVFTAVSTLMENGVPSPATPENILKEATHVTSQGYEDKSDWMBIGWIYG	760	QY	701 MSLEKRIGQSAVFTAVSTLMENGVPSPATPENILKEATHVTSQGYEDKSDWMBIGWIYG
Db	715 MSLEKRIGQSAFVASTLMEYGGVPOSATPESLIKAEATHVTSQGYEDKREWTEIGWIYG	774	QY	715 MSLEKRIGQSAFVASTLMEYGGVPOSATPESLIKAEATHVTSQGYEDKREWTEIGWIYG
QY	761 SVTEDDLTGKHKHARGRSYIYCMPLKAFKGSPAINUSDRNLQYRWAALGSVETLFSHC	820	QY	761 SVTEDDLTGKHKHARGRSYIYCMPLKAFKGSPAINUSDRNLQYRWAALGSVETLFSHC
Db	775 SVTEDDLTGKHKHARGRSYIYCMPLKAFKGSPAINUSDRNLQYRWAALGSVETLFSHC	834	Db	775 SVTEDDLTGKHKHARGRSYIYCMPLKAFKGSPAINUSDRNLQYRWAALGSVETLFSHC
QY	821 PIWGYKWLKFELRFAYVNTTIPPTISIPLMCTLLAVCLFTNQFIPQSNIASTWP	880	QY	821 PIWGYKWLKFELRFAYVNTTIPPTISIPLMCTLLAVCLFTNQFIPQSNIASTWP
Db	: :	880	Db	: :
QY	835 PLWYGYGGRKLKFELRFAYINTTIPPLSPLPLLYCILPAICLTLGKFPIEISNFASTWP	894	QY	835 PLWYGYGGRKLKFELRFAYINTTIPPLSPLPLLYCILPAICLTLGKFPIEISNFASTWP
Db	881 LSLFLSIFATGPLEMMRNSGVDEDEWRNEQFWYIGVVAHLFAVEQGIIKVLAGIDNTNT	940	Db	881 LSLFLSIFATGPLEMMRNSGVDEDEWRNEQFWYIGVVAHLFAVEQGIIKVLAGIDNTNT
QY	: :	940	QY	: :
Db	895 ISLFSISFATGLEMNRNSGVDEDEWRNEQFWYIGVVAHLFAVEQGIIKVLAGIDNTNT	954	Db	895 ISLFSISFATGLEMNRNSGVDEDEWRNEQFWYIGVVAHLFAVEQGIIKVLAGIDNTNT
QY	941 VTKSASIDEDGDPDAELYKLPWTLLIPPTLTLVNLGVYVAGSYAINGYQSMGPFKKL	1000	QY	941 VTKSASIDEDGDPDAELYKLPWTLLIPPTLTLVNLGVYVAGSYAINGYQSMGPFKKL
Db	955 VTKSASIDEDGDPDAELYKLPWTLLIPPTLTLVNLGVYVAGSYAINGYQSMGPFKKL	1014	Db	955 VTKSASIDEDGDPDAELYKLPWTLLIPPTLTLVNLGVYVAGSYAINGYQSMGPFKKL
QY	1001 FFAFWWVYHPLGPLGMGRQRTPTIVWWSVLLASIFSLWVARRDFTSRVGPDBILE	1060	QY	1001 FFAFWWVYHPLGPLGMGRQRTPTIVWWSVLLASIFSLWVARRDFTSRVGPDBILE
Db	: :	1060	Db	: :
QY	1015 FFAFWWVYHPLGPLGMGRQRTPTIVWWSVLLASIFSLWVARRDFTSRVGPDBILE	1074	QY	1015 FFAFWWVYHPLGPLGMGRQRTPTIVWWSVLLASIFSLWVARRDFTSRVGPDBILE
Db	1061 CGINC 1065	1075 CGINC 1079	Db	1061 CGINC 1065
QY	1075 CGINC 1079		QY	1075 CGINC 1079
RESULT 6	Q9L15		QY	582 DGDIDKNDRYANTNTVFDINJNLGLDGQGPVYVGTGCVENRITALGYEPPKVKHKPSL

DE	DT	01-OCT-2000 (TREMBREL. 15, Last sequence update)	DT	01-JUN-2002 (TREMBREL. 21, Last annotation update)
GN	GN	Cellobiose synthase-2.	GN	
OS	OS	<i>Zea mays</i> (Maize).	OS	
OC	OC	Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	OC	
OC	OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;	OC	
OC	OC	Panicoidae; Andropogoneae; Zea.	OC	
NCBI	NCBI	TaxID=4577;	NCBI	
RN	RN	[1]	RN	
RP	RP	SEQUENCE FROM N.A.	RP	
RX	RX	MEDLINE=20198328; PubMed=10938350;	RX	
RA	RA	Holland N., Holland D., Healentjaris T., Dhuga K.S., Xocostle-Cazares B., Delmer D.P.;	RA	
RT	RT	"A comparative analysis of the plant cellulose synthase (CESA) gene family." <i>Plant Physiol.</i> 123:1313-1324 (2000).	RT	
RL	RL	EMBL; AF200526; AAF83962; 1; -.	RL	
DR	DR	InterPro; IPR05150; ZnF_ring.	DR	
DR	DR	Pfam; PF0352; Cellulose_synt; 1.	DR	
DR	DR	SMART; SM00184; RING_1.	DR	
SQ	SEQUENCE	1074 AA; 121187 MW; A116B66F0564E210 CRC64;	SQ	
Query Match	Best Local Similarity	70.9%; Score 4024; DB 10; Length 1074; Matches 749; Conservative 69.4%; Pred. No. 9e-304; Mismatches 159; Indels 44; Gaps 10;	Query Match	
Qy	2	ESBEGETAGKPMKNTVPOTCQICSDNNVKGTRVGDGFRCVACDICSFCVCPYERYERDGNO	Qy	
Db	23	DSDAPVPAKPTKSANGQVQICSDTGVGSATGDVFVACNECAFFVCRPQEYEREGNQC	Db	
Qy	62	CPOCKTRYKRLKGSPAPGDKDGDLAGDEGTVEENYPOKEKISERMLGHILTRKGGEEMG	Qy	
Db	83	CPOCKTRYKRQKGSPRVHGDDEEBDV-DDDNENFYKO-----GNGKGPENQ	Db	
Qy	122	EQYDKEVS-----HHLPLRITSQDTSGERSAASPERLSSSTIAGKRLPSSDVN	Qy	
Db	129	LGQDDALSSSARHDPHIRIPUTGQQISGPADSPRHST-----RSPISYYD	Db	
Qy	175	QS--PNRRIVDP----VGLGNWAKKERDGWAKMQEKTGPySTQAAERGGVIDAS	Qy	
Db	181	PSVPPVPRIVPSKDLNSYGLNSVDWKERVEWSWRVKQDANMLOQTNKYBARGDME--G	Db	
Qy	227	TDLIADEALLNDEARQPLSRKVSPSSRNPYRMIMRLVILVLLHYRITNPVNAFA	Qy	
Db	238	TGSNGEDMQMDVNDARPLSLRIVPISQNLNLYRIVILRLILLCFFQYRISHPVNAYG	Db	
Qy	287	LMLNSVICEIWFLSWSILDQFPKWPVNRETYLDRLALRYDREGEPSQLAVIDIFVSTD	Qy	
Db	298	LMLNSVICEWFALSWLQDFPKWYPINRETYLDRLLALRYDREGEPSQLAVIDIFVSTD	Db	
Qy	347	PIKEPLVLTANTVLILAVDYPVDKVSKCVVFDDGAGMSFESLAETSEPARKWWFCKK	Qy	
Db	358	PIKEPLVLTANTVLISLAVDYPVDKVSKCVVSDGGSAMLTIFESLSSETAEPAKWWFCKK	Db	
Qy	407	SIPEPPLVLTANTVLISLAVDYPVDKVSKCVVFDDGAGMSFESLAETSEPARKWWFCKK	Qy	
Db	418	PIKEPLVLTANTVLISLAVDYPVDKVSKCVVFDDGAGMSFESLAETSEPARKWWFCKK	Db	
Qy	467	MQGTPWPWGNNTQDHGMIQVFLQGONGGLDAEGNEELPLRVYVSREKRGFOHHKKAGMN	Qy	
Db	478	MADGATMPGNPNRDPHGMQVFLQGSGGIDTGMELPRIVYVSREKRGFOHHKKAGMN	Db	
Qy	527	ALYRSAYVLTINGPFILNLCDHDHTINSKALREAMCFMLDPNLKGQCYVYOPRDGIDR	Qy	
Db	538	ALYRSAYLVTLNGAYLNDCDHFFNSKALREAMCFMLDPNLKGQCYVYOPRDGIDR	Db	
Qy	587	NDRYANRNTVFEEDINLRGDLGTOCPVYVGTCVFNRTALYGYERPPIKVKKPSLISKLC	Qy	
Db	598	NDRYANRNTVFEEDINLRGDLGTOCPVYVGTCVFNRTALYGYERPPIKVKKPSLISKLC	Db	
Db	646	559 HDRYANRNTVFEEDINMKGLDGT1QGPVYVGTCVFNRTALYGYERPPIKVKKPSLISKLC	Db	
Qy	657	560 HDRYANRNTVFEEDINMKGLDGT1QGPVYVGTCVFNRTALYGYERPPIKVKKPSLISKLC	Qy	
Db	647	GSRKRKNSKAKKEDKKSSGRHDTSTVPUFNLDIIEGVSEGAGTEDDEKALLMSOMSLER	Db	
DB	658	CGRRKRNKSYMSQSRIIMKR-TESSADJFNMEDIEGIE--GYEDERSVLMRSURKLER	DB	
Qy	707	FGOSAVFVASTLMEENGYPSSATPENLKEAIHVISCGYEDKSDWGMIGWIGSVTEDI	Qy	
Db	715	FQOSPIFIASTTNTQGQPSPINPASILAHIVSCYEGKEWKGKGVGSVTDI	Db	
Qy	767	LNGPKMHARGWASIVCMRKLPAKGSAPINSLRSNLRNOLVWALGSSVELFSRACPIWG	Qy	
Db	775	LNGPKMHARGWOSIYCMPPRCFKGSAPINSLRSNLRNOLVWALGSSVELFSRACPIWG	Db	
Qy	827	NGLRKLFERFAYNTTIPITSPLMLYCTLAVCLFNFQFILPOISNIAINSIPLSLS	Qy	
Db	835	NGLRKLFERFAYNTTIPITSPLMLYCTLAVCLFNFQFILPOISNIAINSIPLSLS	Db	
Qy	887	IFATGILELMRMSGVGIDEMWRNQFWVIGGSVSAHLFAYFQGILKVLAGIDNFTVSKAS	Qy	
Db	895	IFATGILELMRMSGVGIDEMWRNQFWVIGGSVSAHLFAYFQGILKVLAGIDNFTVSKAS	Db	
Qy	947	DEGGDFABFLYFLKWTLLIPPTMLLIVLNUVWAGVSTAINSGQOSWGLFLGKLFKAFAW	Qy	
Db	955	DEGGDFABFLYFLKWTLLIPPTMLLIVLNUVWAGVSTAINSGQOSWGLFLGKLFKAFAW	Db	
Qy	1007	IWHYLPFLKGAGQRONRPTIVVWSVLAISLWIDPFTSRVIGPFDL-ECGTM	Qy	
Db	1015	IWHYLPFLKGAGQRONRPTIVVWSVLAISLWIDPFTSRVIGPFDL-ECGTM	Db	
RESULT	9		RESULT	9
ID	048945		ID	048945
AC	048946		AC	048946
DT	01-JUN-1998 (TREMBREL. 06, Created)		DT	01-JUN-1998 (TREMBREL. 06, Last sequence update)
DT	01-JUN-2002 (TREMBREL. 21, Last annotation update)		DE	Cellulose synthase catalytic subunit.
GN	RSWI OR F8B4.110 OR At4G32410.		GN	Arabidopsis thaliana (Mouse-ear cress).
OS	Eukaryot, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		OS	Eukaryot, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermato phyta; Magnoliophyta; eudicots; rosids; rosidae;		OC	Spermato phyta; Magnoliophyta; eudicots; rosids; rosidae;
OC	eurosid s II; Brassicales; Brassicaceae; Arabidopsis.		OC	eurosid s II; Brassicales; Brassicaceae; Arabidopsis.
OX	NCBI_TaxID=3702;		OX	NCBI_TaxID=3702;
RN			RN	
RP	SEQUENCE FROM N.A.		RP	SEQUENCE FROM N.A.
RC	STRAIN=CY_COLUMBIAN		RC	STRAIN=CY_COLUMBIAN
MDLINE	-9811412; PubMed=9445479;		MDLINE	-9811412; PubMed=9445479;
RA	Arioli T., Peng L., Betzner A.S., Burn J., Witteke W., Hertz W., Camilleri C., Hofte H., Plazinski J., Birch R., Cork A., Glover J., Redmond J., Williamson R.E.,		RA	Arioli T., Peng L., Betzner A.S., Burn J., Witteke W., Hertz W., Camilleri C., Hofte H., Plazinski J., Birch R., Cork A., Glover J., Redmond J., Williamson R.E.,
RT	"Molecular analysis of cellulose biosynthesis in Arabidopsis."		RT	"Molecular analysis of cellulose biosynthesis in Arabidopsis."
RL	Science 279:717-720(1998).		RL	Science 279:717-720(1998).
RN			RN	
RP	SEQUENCE FROM N.A.		RP	SEQUENCE FROM N.A.
RA	Beyvan M., Terrin N., Ardiles W., Buryshaert C., Dasseville R., De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H., Villarcel R., Villarcel R., Gielen J., Van Montagu M., Hoheisel J., Mewes H.W., Mayer K.F.X., Lemcke K., Schueler C.; Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.		RA	Beyvan M., Terrin N., Ardiles W., Buryshaert C., Dasseville R., De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H., Villarcel R., Villarcel R., Gielen J., Van Montagu M., Hoheisel J., Mewes H.W., Mayer K.F.X., Lemcke K., Schueler C.; Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN			RN	
RP	SEQUENCE FROM N.A.		RP	SEQUENCE FROM N.A.
RA	EU Arabidopsis sequencing project;		RA	EU Arabidopsis sequencing project;
RL	Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.		RL	Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN			RN	
SEQUENCE	FROM N.A.		SEQUENCE	FROM N.A.
RA	Terryn N., Ardiles W., Buyschaert C., Dasseville R., De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H., Villarcel R., Gielen J., Van Montagu M., Mewes H.W., Lemcke K., Mayer K.F.X.,		RA	Terryn N., Ardiles W., Buyschaert C., Dasseville R., De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H., Villarcel R., Gielen J., Van Montagu M., Mewes H.W., Lemcke K., Mayer K.F.X.,
RN			RN	
[5]	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.		[5]	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
SEQUENCE	FROM N.A.		SEQUENCE	FROM N.A.
RA	Arabidopsis sequencing project;		RA	Arabidopsis sequencing project;
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.		RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR	AF02172; AAC39334.1; -.		DR	AF02172; AAC39334.1; -.

Qy	995	PFLGKLEFAFWVIVHLYPFLKGMLGRQNTPTIVWWSVLASIFSLWVRFPIFRVT	1054	Qy	462	EEGVWMQDGTPWPGNNNTGDHPGMIQVFGONGGLDAEGNELPRLVYVSREKRGFFHHK	521
Db	1015	PLFGKLFIAFWVIVHLYPFLKGMLGRQNTPTIVWWSVLASIFSLWVRFPIFRVT	1074	Db	484	EEGVWMQDGTPWPGNNNTGDHPGMIQVFGONGGLDAEGNELPRLVYVSREKRGFFHHK	543
Qy	1055	GPILECGCIN 1065		Qy	522	AGAMNALLRVSAVLTNGFILNIUDCHTNTNSKALREAMCFJMDPNLNGKQVYQOPORF	581
Db	1075	GPLIEECGLDC 1085		Db	544	AGAMNALRVSAVLSNAAYLLNLNDCDHYINNSKAKEAMCFJMDPNLNGKQVYQOPORF	603
RESULT 12							
ID	09LL12	PRELIMINARY;	PRT;	09LL12	SEQUENCE FROM N A.		
AC	09LL12;			AC	09LL12;		
DT	01-OCT-2000	(TREMBrel. 15, Created)		DT	01-OCT-2000	(TREMBrel. 15, Last sequence update)	
DT	01-JUN-2002	(TREMBrel. 21, Last annotation update)		DE	Cellulose synthase-8.		
GN	CESA-8.			OS	zea mays (maize).		
	Eukaryota; Viridiplantae; Streptophytta; Embryophytta; Tracheophytta;			OC	Spermatophytta; Magnoliophytta; Liliopsida; Poales; Poaceae; PACC clade;		
OC	Panicidae; Andropogoneae; Zea.			OX	NCBI_TaxID:4577;		
RN	[1]			RP	SEQUENCE FROM N A.		
RX	MEDLINE=200398328; PubMed=10938350;			RX	Holland N., Holland D., Heleentjaris T., Dhugga K.S.,		
RA	Xocomostle-Cazares B., Delmer D.P.;			RA	"A comparative analysis of the plant cellulose synthase (CESA) gene family.";		
RT	Plant Physiol. 123:1313-1324 (2000).			RL	PLANT PHYSIOL. 123:1313-1324 (2000).		
DR	AER00532; AAC89968_1;			DR	InterPro; IPR00150; Cellulose_synt.		
DR	InterPro; IPR00184_1; Znf_ring.			DR	pfam; PF0552; cellulose_synt; 1.		
DR	SMART; SM00184; RING_1.			DR	SMART; SM00184; RING_1; 122575 MW; DBBCC18F49D23F5E CRC64;		
SQ	2	RSBGETAGPKMKNIPQPCSDNWRKTVDDEFVACDICEFPVCAPCYEERKQDQS	61	Qy	810	GSVEILFSRHCPIWYGNGRKLFRAYVNNTIYPISTIPLMCTILAVLIFTNOFI	869
Db	DRESGAAGGGAAARRAEPACOICGDEVGVGFDFSEPPVACNECAFPCVACAYEYERREGSQ	82		Qy	839	GSYEVTSKHKCPIWYGNGRKLFRAYVNNTIYPISTIPLMCTILAVLIFTNOFI	898
Oy	62	CPOCKTRKRLKGSPAPLGKDGLAD-----EGTVEENPVQEKEISERMGLGHMT	113	Qy	870	PQTSNIASTIWLSLFLSIFATGILEMRMSGVGDEWHNRNEQFWIGVSAHLFAVRGIL	929
Db	83	CPOCRTRKRLKGSPAPLGKDGLAD-----EGTVEENPVQEKEISERMGLGHMT	113	Qy	899	PELTINVASIWFMALFCISVTGIGLMEWRMSGVAIDWWNRNEQFWIGVSAHLFAVRGIL	958
Oy	114	RSGKGEEMCEPQDKEVSHNHLRPLTSRO--DTSGEASAASPERLSSSTSITAGKR	167	Db	930	KVLAGIDDNFTWISKASIDEGPAAELFKWTTLIPPTTLIUVNLYGVWAGVSYAINTSG	989
Db	141	:----- :----- :----- :----- :----- :----- :-----	194	Qy	959	KVFRAGIDSFTWISKAGD-DEESETLTFTWKLPTTLLIPPTTLIUVNLYGVWAGVSYAINTSG	1017
Oy	168	PISSDVNUQSPNRRIVD-----VGLCNVAWERDVQMKOBKNTGPVSTOAASERGV	221	Db	990	YQSGWPLFGKLEFAFWVIVHLYPFLKGMLGRQNTPTIVWWSVLASIFSLWVRFPI	1049
Db	195	PF-ADPNLPVQQRFSMDSFSKDLAAYGKGSVAKWERMGEWKQKBR-----LOHRVSEGG	247	Qy	1018	YEWGWLPGKLEFAFWVIVHLYPFLKGMLGRQNTPTIVWWSVLASIFSLWVRFPI	1077
Oy	222	DIDASTDILADRLANDEAROPLSKRSVIPSRRNPYRMVMLRVTLCLFLHYRINTPV	281	Db	1050	TSRVTGPPILECGCIN 1065	
Db	248	DWDGDD--ADLPLM-DEARPLSKRVPISSRNPYRMVTLRVVLFHYRVMHP	303	Qy	1078	LAKSNGLPPILECGLDC 1093	
RESULT 13							
ID	09SWW6	PRELIMINARY;	PRT;	09SWW6	SEQUENCE FROM N A.		
AC	09SWW6;			AC	09SWW6;		
DT	01-MAY-2000	(TREMBrel. 13, Created)		DT	01-MAY-2000	(TREMBrel. 13, Last sequence update)	
DT	01-MAY-2002	(TREMBrel. 21, Last annotation update)		DE	Cellulose synthase catalytic subunit.		
GN	ITX3 OR T10B6-80.			OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophytta; Tracheophytta;			OC	Spermatophytta; Magnoliophytta; eudicotyledons; core eudicots; Rosidae;		
OC	Eurosidii II; Brassicales; Brassicaceae; Arabidopsis.			OC	NOB1_TaxID:3702;		
RN	[1]			RN	SEQUENCE FROM N A.		
RP	SEQUENCE FROM N A.			RC	STRAIN=CV. LANDSBERG ERECTA;		
RX	STRAIN=CV. LANDSBERG ERECTA;			RX	MEDLINE=9924300; PubMed=1030464;		
Db	304	KDFAWLILSWSVCEIWPMWSLDFPKWLPITERVYDRLSLRDFREGQPSQALDF	363	RT	Taylor N.G., Scheible W.R., Cutler S., Somerville C.R., Turner S.R.;		
Qy	342	VSTVDPDKEPKPLVANTVSLILADVYDVKVACVYFDGAMSFESLAESEPARGV	401	RT	"The irregular xylem locus of Arabidopsis encodes a cellulose synthase required for secondary cell wall synthesis.";		
Db	364	VSTVDPDKEPKPLVANTVSLILADVYDVKVACVYFDGAMSFESLAESEPARGV	423	RT	plant Cell 11:769-780(1999).		
Qy	402	FCKKYSTEPRAPEWYFAKIDLKDVYQTVFVYKDRRAMKREVEEKKRINALVAKJC	461	RN	SEQUENCE FROM N A.		
Db	424	RAFCKFNIEPRAPEWYFOOKIDLKDVYQTVFVYKDRRAMKREVEEKKRINALVAKJC	483	RP	Sato S., Nakamura Y., Kaneko T., Kato T., Asanizu E., Kotani H., Tabata S., Mewes H.W., Rudd S., Lemke K., Mayer K.F.X.; Submitted (AUG-2000) to the EMBL/GenBank/DDJB databases.		

QY	356 ANTVLISLAVDYPDKVSCYYEDGAAMSFESLAESSEFARKWMPCKYSIESTPAREW 415	RA	Salzberg S.L., Fraser C.M., Venter J.C.;
Db	335 SNTVLISLAMDYPEVKTSVSDGASMLTFESLSETABFKWNPKFKESTEPAPEM 394	RT	"Sequence and analysis of chromosome 2 of the plant <i>Arabidopsis</i> thaliana.".
QY	416 YFAAKIDYKDKQTSFWDKRAMKREYEFKIRINALVSKALKCPERGVMDGTWPG 475	RL	Nature 402:761-768(1999).
Db	395 YFTLKVDYQDKYHPTFVYKERRAMKREYEFKIRINALVSKALKCPERGVMDGTWPG 454	RN	[2]
QY	36 TNEPFPILNDCDIYINNERALRERACFLMDPNIGKOVYVOFOREPGIDKNDYRANT 595	RP	SEQUENCE FROM N.A.
Db	515 TNAFPMLNLDCDHVNNSKAVREACFLMDPQIKKVYVQFFORFDGIDTNDYRANT 574	RC	S ^{TRAIN} -CV, COLUMBIA;
QY	596 VFPDINLQLGDLGTOGPVWGTGVCFNRPALYGEPEPIKWKPPSLSLKCG---GSR 650	RA	Lin X.;
Db	575 VFPDINLKGKLDGIOGPVWGTGVCFNRPALYGEPEPIKWKPPSLSLKCG---GSR 630	RA	Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
QY	651 KKNKAKKESDKKKSGRHTSTWPVFNLDIEBVEG AGFDDKEKALIMQSLEKREGQ 709	DR	EMBL: AC006300; AAC2013_1; -.
Db	631 KRKKK-----FSRNMDNGDVAALGGAEDKDEHLMFEMFEKTFGQ 670	DR	InterPro: IPR05150; Cellulose_synt.
QY	710 SAVFVASTMENGGVPPATPENILKEAHLVSCGYEDKSNDMEIGHYGSPEDELIG 769	DR	InterPro: IPR01841; Znf_ring.
Db	671 SSIFVTSILMEEGVPPSSPAVLIKAEHVSCGYEDKSNDMEIGHYGSPEDELIG 730	PFAM	PF0352; Cellulose_synt; 1.
QY	770 FKMARGMSIYCEPKLAKFGSAPINISDRNLNOVLRVALGSEVILSRHCPIWGY -NG 828	DR	InterPro: IPR01841; RING; 1.
Db	731 FKMHCGRWHSIYCEPKRPAFKGSAPINLSDRLNQVLWALGSVEIFFRSRHSPLWGYKG 790	PROSITE	PS01121; CASPASE_HIS; UNKNOWN_1.
QY	829 RUKFLERFAYNTITPISIPLMCLLAVLCLFTNOEIIPOQNSNFSIWFSLFUSF 888	DR	SEQUENCE: 1065 AA; 120616 MW; F2AE36C9492A6076 CRC64;
Db	791 KKLWLERFAYNTITPISIPLMCLLAVLCLFTNOEIIPOQNSNFSIWFSLFUSF 850	Query Match	67.0%; Best Local Similarity 66.2%; Pred. No. 2e-286; Length 1065; Matches 714; Conservative 132; Mismatches 156; Indels 76; Gaps 12;
QY	949 DGFDAELYLFKWTLLIPPTTLIVNULVGVVAGVSYAINSGQSWGRPLFGKLFPAWIV 1008	Db	10 KPMKNNIVPQTQCQCSDNVKGKTVGDGRFVACDCSFVCPRCYEVREDGQNSCPOQTRY 69
Db	911 D-DFGELIAKFWTLLIPPTTLIVNULVGVVAGVSYAINSGQSWGRPLFGKLFPAWIV 969	QY	11 KPLKDNLNQICQICGDDVGLTKTGNVFCACNCBCGFLQSCSYEYERDGSOCCPOCKARF 82
QY	1009 HLYPFLKGIMGRONRTPPTVWWSVLLASIFSJLWWRDPFISRVGPDILEGGINC 1065	Db	12 KRUKGSPAIPGDKDEGDADEGVFENPOKEKISERMLGWLTRGRGEEMCEPQDKEV 129
Db	970 HLYPFLKGIMGRONRTPPTVWWSVLLASIFSJLWWRDPFISRVGPDILEGGINC 1026	QY	13 83 RRHNGSPREVDSKEDWDVNDIEN-EFDVQ----- 111
RESULT 15			
Q9SKJ5	PRELIMINARY; PRT; 1065 AA.	Db	14 DQQLPGIYQQLLAPVRILDPSKDLNQSYLVNVWDWKRIQWKKRQDKNMHRGTGHHEGK 223
ID	Q9SKJ5	QY	15 219 GGVIDIDASTDILADEALINDEARQPLSKRKVSPSSRNPYRVMILRVLFLFHRYIT 278
AC	Q9SKJ5	Db	16 224 GGE--FEGTGNSGDELQDMLVADPLRMSVPMVWHPSPARMPYRIVLRLIIGLFLHYRT 281
DT	01-MAY-2000 (TREMBrel. 13, Last sequence update)	QY	17 279 NPVPNAFWLVCICEWFASWLDQOPKPWNPVNRNTYLRALRDREEPSQAAV 338
DT	01-JUN 2002 (TREMBrel. 21, Last annotation update)	Db	18 282 HPKDAYALWLSVCIWFASFWLQDOPPKWNPVNRNTYLRALRDREEPSQAAV 341
DE	putative cellulose synthase catalytic subunit.	QY	19 339 DIVSTVPLKEPLVANTVLSILAVDYPVVKVSCYFDDGAMMSFESLAESTEPARK 398
GN	AT2G25440	Db	20 342 DVFWSTVPLKEPLVANTVLSILAVDYPVVKVSCYFDDGAMMSFESLAESTEPARK 401
OS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; OC	QY	21 399 WPFCKYSIESTPAREWYFAAKDYLKDQVQTSFKRARRMREYEFKIRNALVSKAL 458
OC	Spermatophytia; Magnoliophyta; eu dicots; core eudicots; Rosidae; OX	Db	22 402 WPFCKKENIEPRAPEFYSOKIDYKDKIOPSFVKERRMREYEFKIRNALVSKAL 461
NCBI_TaxID	3702;	QY	23 459 KCPEBGMQDGIPWPGNNTGDPGMLOVFLGONGAAGELPFLRPLVYSEKFRVH 518
RP	SEQUENCE FROM N.A.	Db	24 462 KIFEGWTMEDGTSWPGNPRDHPGMLOVFLGONGAAGELPFLRPLVYSEKFRVH 521
RC	STRAIN-CV, COLUMBIA;	QY	25 519 HKRAGAMMALVRYSAVLTNGPTEFLNLDCHYTNNSKALREAMCFLMDPNLGQVCKVQFP 578
RA	MEDLINE=20033487; Published=10617197;	Db	26 522 HKRAGAMMALVRYSAVLTNGPTEFLNLDCHYTNNSKALREAMCFLMDPAIGKRCYVQFP 581
RA	Lin X., Kaul S., Rounseley S.D., Shea T.P., Benito M.-I., Town C.D., Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V., Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S., Cronin L.A., Shen M., Vanaken A., Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R., Copenhafer G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,	QY	27 579 QRDGDKNDYRANRNTVFFDINLRGLDGIOPVWGTGVCFNRALYGEPEPIKWK 638
RA	Copenhafer G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,	Db	28 582 QRDGDLHDYRANRNTVFFDINLRGLDGIOPVWGTGVCFNRALYGEPLTDELDE 641
RA	Openheimer G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,	QY	29 639 PSULSKLUGGSRRKNSRKA-KESDKRSGRHFSTVVFNLDDIESERVEGAKFDDRALL 697
RA	Openheimer G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,	Db	30 642 PNIVKSCFGSRKKGSKRKIPNEDNNSIKRSDSNPNLFNMEDIDEVE- GYEDMSLL 699
RA	Openheimer G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,	QY	31 698 MSQMSLRERFGGSAVYASTLNGGVPSSAPENILKEAHLVSCGYEDKSDWGREGW 757
RA	Openheimer G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,	Db	32 700 VSQKLERKFGQSPVFIATFMEBOGLGPLSTNTPLTLIKEATHVISGYEAKTDWGREGW 759

QY 758 IYGSYTEDILTGFKMHARGRSIVCAMPKPAFKGSAPINTSDRINQVRLVALGSVEILS 817
Db 760 IYGSYTEDILTGFKMHARGRSIVCVPSPRAFKGSAPINTSDRINQVRLVALGSIEILS 819
QY 818 RHCPIWYGNGRKLTERAYVNTTYPPTSIPLIMCYTLAVCLFTNQTIPOSNIAS 877
Db 820 RHCPIWYGNGRKLTERAYVNTTYPPTSIPLIMLPACFLNTTIPESNLAS 879
Qy 878 IWPLSLFLSFRATGLEMMSGVGDEWMARNEQFWIGYSALHAFVFGTILKVLAGIDT 937
Db 880 LCFMLFASLASYASALEKNSDALEDMARNEQFWIGTSASAHLFAVEFGULKEAGIDT 939
Qy 938 NFTVSKASDEDGDFAEYLVLPKWTULLIPPTLLIVNLGVWVGVSATNSGYOSWGLPF 997
Db 940 NFTVSKASDEDGDFAEYLVFKWTSLLPPTTLLIVNLGVGIVGVSATNSGYOSWGPAM 999
Qy 998 GKLEFAFWWVLYPHLYPLKGIMGROHTPTVVMSVLASIFSLWVRDPFTSEVTG 1055
Db 1000 GKLFAFWWVAHLYPLKGIMGROHTPTVVMSVLASIFSLWVRDPFTSEVTG 1056
1056

Search completed: June 16, 2003, 11:07:59
Job time : 55 secs

THIS PAGE BLANK (USPTO)

(USPTO)